Validation of the Safety Behaviour Test (SBT): Criterion-Related Validity Evidence

---

A dissertation submitted in partial fulfilment of the requirements for

the Degree of

Master of Science

In Applied Psychology

At the University of Canterbury

By Lydia Crowe

2018

Research Supervisors

Associate Professor Christopher Burt, University of Canterbury

Dr Katharina Näswell, University of Canterbury

Table of Contents

List of Tables

List of Figures

Acknowledgements

Firstly, I would like to thank all the study participants. This type of research is not possible without people willing to share their time and experiences.

To Chris Burt, my research supervisor, I'd like to say . . . "Thank you for your ongoing support, time and guidance. I have really appreciated your help with planning and organising my project. Without this and your patience and encouragement, this dissertation would have been far harder to complete."

Lastly, I would like to thank my partner, family, and friends. These people have quietly (and at times, not so quietly) pushed me to deliver my best. In particular, my parents, who without doubt have supported me to do this. For those of you that have challenged my ideas, and proofread my words, I am grateful that you have been there for me.

Abstract

High workplace accident and fatality rates have made the measurement of safety behaviours in job applicants a primary concern for organisations. Employee safety behaviours have been defined by a number of actions, including: safety compliance, participation, voicing, and consciousness. The significant role that these safety behaviours play in maintaining workplace safety argues strongly for the need of tools which can be used to measure these behavioural tendencies in job applicants. With this information, job applicants with these safe behaviours can be selected for high-risk positions over those without. To address this problem the Safety Behaviour Test (SBT) was developed as a gamified assessment tool, designed to objectively measure safety behaviour within an animated work environment simulation. The main purpose of this study was to investigate whether the SBT has criterion-related validity. To conduct the validation 200 participants were recruited. 100 of which completed the SBT (SBT participants), while the other 100 participated as independent criterion data sources (acquaintances) and reported on the SBT participants' safety behaviours. The SBT participants' scores were correlated with the data on the individual's actual safety behaviour provided by the acquaintance. Results indicated that the SBT captures authentic safety behaviour and has criterion-related validity. The practical and theoretical value of the SBT's criterion-related validity in the reduction of workplace accidents is discussed.

Overview

The introduction begins by outlining New Zealand's health and safety context, and considers the idea that New Zealand's high workplace accident rates may be the result of poor employee safety behaviour decisions. Health and safety research models were investigated and provided further evidence that employee safety behaviour decisions have the greatest impact on workplace accidents.

Given the primary influence of safety behaviour decisions in accident prevention, measurement of these behaviours before employment is recognised as essential in accident prevention. To establish if this type of measurement is currently available, the introduction reviews health and safety psychometric measures and concludes that they do not accurately measure safety behaviours. These psychometrics often measure less influential factors in workplace accidents through a self-report medium that is susceptible to social desirability and impression management biases.

The capability to test through a game-based format was investigated, with gamified tests' ability to potentially overcome impression management bias and increase measurement accuracy discussed. As a gamified assessment, the SBT is then proposed as a solution to the need for an unbiased measure of safety behaviours. The description of the SBT is followed by information regarding research's mixed reports for gamified test validity and the need for the SBT to be validated before it can be considered as a solution. Finally, how the current study investigated and answered its research question 'Does the SBT have criterion-related validity?' is explained.

Introduction

The Extent of Workplace Accidents

Worldwide the workplace is responsible for one death and 153 work-related accidents every 15 seconds, which results in an estimated 2.3 million deaths and 313 million non-fatal accidents annually (International Labour Office, 2017). Unfortunately the New Zealand work environment has contributed to these concerning global statistics. Although total accident and fatality figures in New Zealand have slowly declined in the last three years, total recorded work-related injuries remain at a level of major concern (Worksafe NZ, 2016). Specifically, between June 2015 and June 2016 New Zealand workplace accidents were responsible for 49 fatalities and 3,384 serious non-fatal injuries (Worksafe NZ, 2016). The cost of these accidents on human lives is evident in the high national and global fatality rates (International Labour Office, 2017; Worksafe NZ, 2016). However, the impact of workplace accidents exceeds an employee's physical pain, impacting on the injured employee's family, organisation, and wider society (ACC, 2016; Hofmann, Morgeson & Gerras, 2003; International Labour Office, 2017).

Workplace accidents impact the injured employee through medical visit costs, family distress, losses related to lifestyle changes, and income (Worksafe NZ, 2016). The organisation can experience significant damage to their reputation for social performance (Barney, 1991; Orlitzky & Benjamin, 2001); this can limit an organisation's competitive advantage (Barney, 1991), recruiting attempts (Hofmann *et al*, 2003), morale, and production, while increasing employee recruitment and training costs (Worksafe NZ, 2016). On a societal level workplace accidents are responsible for significant economic costs. Figures from New Zealand's Accident Compensation Corporation reveal that close to 700 million dollars was paid out for 303,000 workplace injuries in a 12 month period ending in June 2015 (Accident Compensation Corporation, 2016). Given the extensive impact of these workplace accidents, analysis of contributing causes is of great value.

New Zealand workplace statistics offer some initial context surrounding potential causes of workplace accidents. New Zealand statistics, for example, show those with the highest risk for workplace accidents are new employees with short job tenure (Burt, 2015;

Worksafe NZ, 2016). Of the 49 fatalities and 3,384 serious non-fatal workplace accidents, which occurred in New Zealand in between June 2015 and 2016, new employees had a disproportionately high representation (Worksafe NZ, 2016). A number of researchers have examined the relationship between job tenure and employee accidents and reported employee job tenure's negative relationship to workplace accidents as a common and global trend (Breslin & Smith, 2006; Burt, 2015; McCall & Horwitz, 2005; Morassaei, Breslin, Shen, & Smith, 2013; Root & Hoefer, 1979). For example, McCall and Horwitz (2005) examined 1,168 trucking vehicular accidents and reported that 51% of these accidents were made by drivers with less than one year of job tenure. Root and Hoefer (1979) examined the data of approximately 270,000 workplace accidents from ten US states and found (regardless of industry, age, or sex) that the highest percentage of these accidents occurred during the first year of employment (40%), with half of these occurring during an employee's first three months on the job.

New Zealand statistics provide some insight into the causes of these workplace accidents, identifying both unsafe situational environments and employee behavioural decisions. Specifically, revealing employee behavioural decisions as having the greatest influence over that of situational factors. Of the 49 work-related fatalities that occurred in 2016 within New Zealand, only 13 fatalities appear to be the result of workplace situational influences; while 39 were primarily the result of human behavioural decisions (Worksafe NZ, 2016). Workplace safety models provide greater insight into how these workplace situational factors and employee behavioural decisions have contributed to New Zealand's high workplace fatality rates.

Models of Safety

Reason's (1990) Swiss Cheese Model explains how New Zealand's workplace situational factors contribute to its high workplace accident rates through the creation of hazardous situations. If situational factors such as organisational supervision, environment, equipment, and operating conditions are *unsafe* they will create a hazardous situation (Reason, 1990). For example, consider one of New Zealand's 2016 workplace fatalities which occurred when a truck driver reversed over a motorcyclist who had followed it onto a worksite (Safeguard, 2017). In this workplace accident the hazardous situation was the motorcyclist on the worksite. Three unsafe factors created this hazardous situation, including unsafe operating conditions **(**with no site-specific hazard register and, as a result, no preparation for vehicles following trucks past the stop signs), unsafe equipment (with broken reverse lights and a broken reverse warning device), and unsafe supervision (with only one employee inducted onto the site) (Safeguard, 2017).

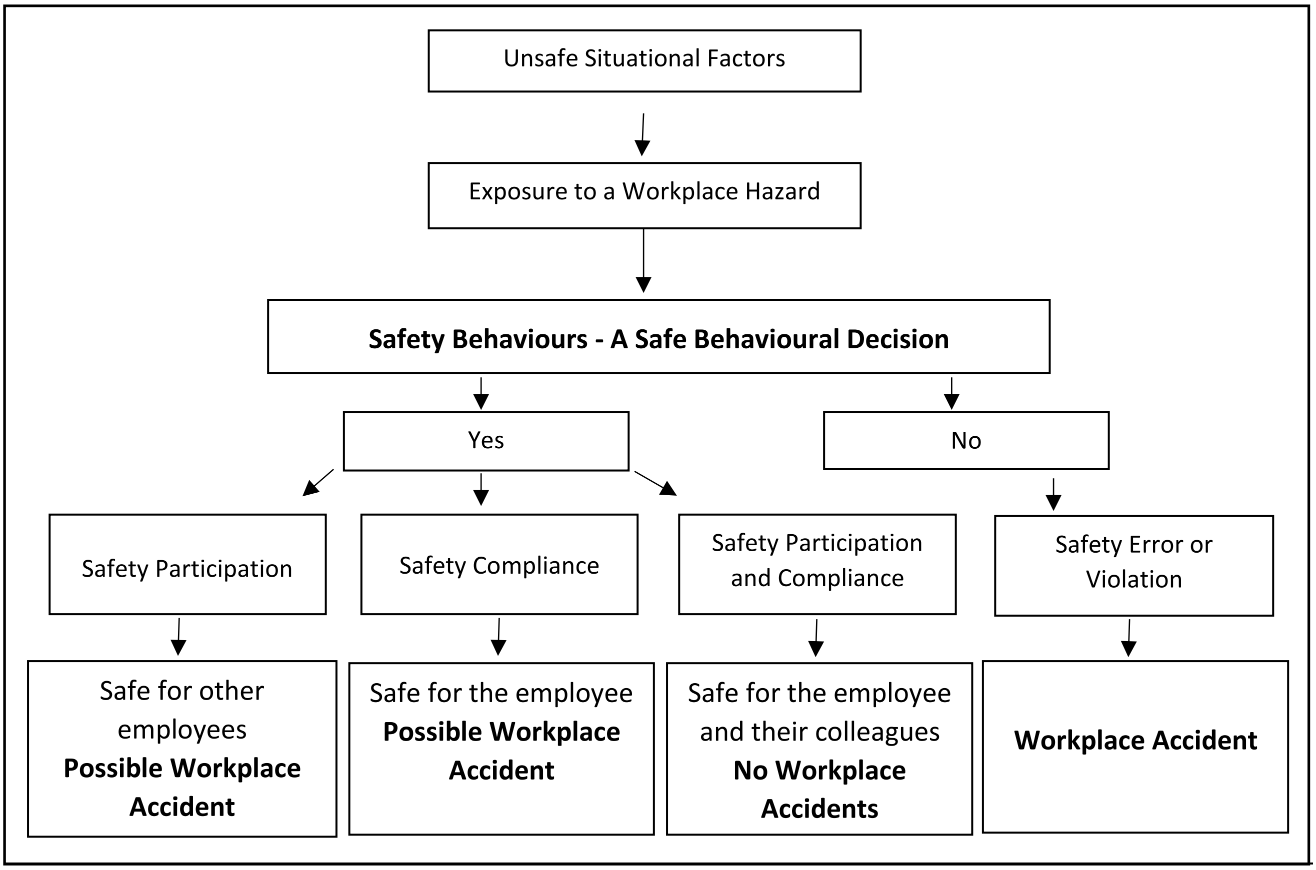
In accordance to Reason's (1990) model, for a hazardous situation like this to result in a workplace accident, an individual *must* make an unsafe behavioural decision, a safety error (skill, decision, or perceptual based) or safety violation (routine or exceptional behavioural violations of safe company practice), in response to the hazardous situation (see Figure 1). These unsafe behavioural actions *entirely mediate* whether or not a workplace hazard results in a workplace accident (see Figure 1).

When applied to the New Zealand example above, the hazardous situation of the motorcyclist on the worksite lead to a workplace fatality due to the safety violation behaviours of the truck driver, who reversed the truck without a spotter (as is company policy). Employers can attempt to minimise workplace accidents through interventions to the work situational environment (Reason, 1990). However, in the end, it comes down to the behavioural actions of employees which dictate whether or not a workplace accident will occur (see Figure 1) (Ramsey, 1985; Reason, 1990; Christian, Bradley, Wallace, & Burke, 2009). Thus, in order to reduce New Zealand's high workplace accident rates safety behaviours which can prevent workplace accidents must be identified.

Ramsey's (1985) Accident Sequence Model identifies a *behavioural decision to avoid* a hazardous situation *must* be made in order to prevent a workplace accident (see Figure 1) (Ramsey, 1985; 1989). However, in order to make this behavioural decision the hazardous situation must be perceived (Ramsey, 1985; 1989). Thus, although this safety behaviour can prevent many workplace accidents it cannot prevent all, as hazards are not always perceived. Consider, for example, the case of the New Zealand workplace fatality discussed above. In these situations safety behaviours which do not require hazard perception are required for accident prevention.

Figure 1

Employee interaction processes with a workplace hazard and the resulting workplace accident or accident avoidance.



Note. Adapted from Reason's (1990) Swiss Cheese Model, Ramsey's (1985) Accident Sequence Model, and Christian *et al*'s (2009) Integrative Model of Workplace Safety.

Christian *et al*’s (2009) Integrative Model of Workplace Safety provides insight into safety behaviours which prevent workplace accidents, both with and without hazard perception. Christian *et al* (2009) recognised two behavioural decisions, safety compliance and safety participation, in accident prevention (see Figure 1) (Christian *et al*, 2009). Safety compliance is ‘generally mandated’ safety behaviours, and refer to an individual’s decision to behave in a manner which will avoid an accident, such as adhering to workplace safety procedures and carrying out one’s work in a safe way (Neal & Griffin, 2006; Neal, Griffin, & Hart, 2000). These compliant behaviours do not require hazard perception and prevent the individual engaging in them from being involved in a workplace accident (see Figure 1). Behaviourally, safety compliance is represented by following procedures, using protective equipment, and practising risk reduction in accordance with company rules (Christian *et al*, 2009; Neal & Griffin, 2004).

Safety participation behaviours are aimed at accident prevention and risk reduction. They prevent colleagues, of the employee engaging in safety participation behaviours, from being involved in a workplace accident (see Figure 1) (Neal & Griffin, 2006). Safety participation behaviours support co-worker safety by promoting the safety program within the workplace, demonstrating initiative, and putting effort into improving safety in the workplace (Neal, Griffin, & Hart, 2000). Behaviourally, safety participation is represented by communication, voice, helping, stewardship, exercising rights, whistleblowing, civic virtue, and initiating safety related change (Christian *et al*, 2009).

In addition to the three safety behaviours discussed are other safety behaviours which can prevent workplace accidents. Two of the most relevant include safety voicing and safety consciousness (Tucker, Chmiel, Turner, Hershcovis, & Stride, 2008; Westaby & Lee, 2003). Safety voicing is “communications directed at improving safety conditions” (Tucker *et al*, 2008, p.319), which reduce workplace accidents by alerting and drawing attention to potential hazards. Safety consciousness is "a positive attitude and awareness toward acting safely" (Westaby & Lee, 2003, p. 228), reducing workplace accidents by enacting behaviours that foster operational safety (de Koster, Stam, & Balk, 2011). These two safety behaviours are two of the most relevant safety behaviours because they overlap, to an extent, with safety participation. As you can see in the definition of safety participation, it includes components of safety voicing and safety consciousness by 'promoting the safety program' and 'voicing' safety concerns. This overlap is typical of safety behaviours.

Research is consistent in the fact that safety behaviours, in particular safety compliance and participation behaviours, hold the greatest influence on workplace accident prevention and causality (Christian *et al,* 2009; Ramsey, 1985; Reason, 1990). Thus, measurement of these and the associated behaviours of safety voicing and consciousness is of significant value in accident prevention (see Figure 1). Employers can improve their company's safety through measurement of these safety behaviours before employment, during recruitment. With this early safety behaviour information organisations can prevent workplace accidents through training or by removing applicants from the recruitment process. Should enough organisations participate, there is the potential to reduce New Zealand's high workplace accident rates. In response to this need, many health and safety psychometric measures have been developed (see Table 1).

Current Psychometric Safety Measures

Upon investigation, it is evident that currently available safety psychometric measures (see Table 1) do not adequately address the need for measurement of safety behaviours during employee selection. Current psychometrics' validity in measuring and predicting safety performance is limited by two key components: their measurement of less predictive factors in workplace accidents and the self-report mode through which they measure.

Currently available psychometric measures attempt to predict employee workplace safety through measurement of factors with limited predictive power. Safety psychometrics (see Table 1) commonly predict safety performance through the measurement of personality, safety knowledge, hazard awareness, and safety motivation, which all have a small influence in the reduction of workplace accidents (Barrett, 2010). Specifically, correlation analysis reveal their influence on the reduction of workplace accidents ranges from only $r = -.11$ to $r = -.21$ (Christian *et al*, 2009). Moreover, the small relationship of these commonly used predictors' in reducing workplace accidents is *mediated* by employee safety behaviours, such as those of safety participation and compliance (Christian *et al*, 2009).

As an example, consider the OPRA Consulting Group's psychometric measure the Health and Safety Indicator (HSI) (see Table 1) (Barrett, 2010; OPRA group, 2017). The HSI psychometric tool includes a combination of targeted personality (safety motivation, diligence, confidence, composure, adherence to rules, and openness to guidance) and safety knowledge measures (understanding instructions, the safety environment, and attention to detail) (Barrett, 2010; OPRA group, 2017). Through the measurement of these personality and safety knowledge factors, the HSI claims to identify/predict an individual's health and safety performance (OPRA group, 2017). However, according to extensive health and safety research, employee personality and safety knowledge have a small influence on reducing workplace accidents, $r = -.21$ and $r = -.11$ respectively (Christian, 2009), which is mediated by safety behaviours (Reason, 1990). An individual can have personality and safety knowledge that lends them to recognising hazardous situations. However, recognition of a hazardous situations does not necessarily translate into safety behaviour decisions which prevent workplace accidents (see Figure 1) (Christian *et al*, 2009; Ramsey, 1985; Reason, 1990).

Table 1

Currently Available Occupational Safety Psychometric Measures and their Publisher

Publisher	Commercial Product
Bay State Psychological Associates Inc.	Employee Reliability Inventory
Hogan Assessment Systems Inc.	Hogan Safe System
IPAT Inc.	Personnel Reaction Blank
OneTest Pty Ltd.	Onetest Work Safety Assessment (OWSA)
Orion System Inc.	Orion Pre Employment System PE3-SAFE
Psyfactors Pty Ltd.	Situational Safety Awareness Test
Psych Press	Work Safety Assessment
RightPeople	RMP Safety Inventory
SHL Plc.	Workplace Safety Solution Test
Synergy Safety Systems	Safety Attitude Survey
Vangent (Pearson) Inc.	Employee Safety Inventory (ESI)
Vangent (Pearson) Inc.	Personnel Selection Inventory (PSI)
OPRA Consulting Group	Health and Safety Indicator (HSI)

Note. Table adapted from Paul Barrett's review of commercial products associated with the psychometric assessment of safety attributes within prospective employees.

The HSI and the psychometric measures listed in Table 1 are also reflective of the majority of available safety psychometric measures in their measurement of health and safety through self-report items (Barrett, 2010). Self-report measures are limited by the potential for their validity to be compromised by common biases such as impression management and social-desirability (Alliger, Lilienfeld, & Mitchell, 1995). When engaging in social desirability, individuals regulate or adjust their answers within measures of traits or behaviours to avoid criticism, for social approval, or to establish a positive impression (Fastame & Penna, 2012; Johnson & Fendrich, 2005; Paulhus, 1984). Impression management can be likened to the concept of social desirability and refers to a goal-directed deception process to create favourable impressions (Schlenker, 1980). Here individuals change, shape, manage, or regulate their answers in order to influence others' perception of them (DuBrin, 2010), in the hope that the impression they form will be a positive one (Fastame & Penna, 2012).

Social desirability and impression management response biases affect the validity of self-report psychometrics (Huang, Liao, & Chang, 1998; Podsakoff, Mackenzie, & Lee, 2003). According to Nederhof (1985), between 10% and 75% of the variance in participants' responses on self-report measures can be explained by these common biases. In research work these common biases can confound relationships among variables measured, either by obscuring variable relationships (King and Brunner, 2000; Van de Mortel, 2008) or by creating artificial relationships which are neither valid nor reliable (Podaskoff, MacKenzie & Lee, 2003). In a practical setting these biases may see an organisation assume an individual will behave in a safe way – when in fact they may not.

For an individual to engage in social desirability or impression management the following three conditions must be met: (1) the measurement context motivates an individual to present themselves favourably, (2) the construct measured has a socially desired response (Furnham, 1986; Villanova & Bernardin, 1991), and (3) the construct intended for measurement is evident to the individual (Furnham, 1986). All currently available safety psychometric measures listed in Table 1 appear to meet these conditions and therefore are susceptible to these common biases. The high-stakes selection context in which these safety psychometric measures are used motivates an individual to present themselves favourably (Bass, 1957). In the measurement of safety behaviours there is, and always will be, a clear socially desirable answer of being safe. Finally, the transparent nature of these self-report psychometrics enables the motivated applicant to engage in these biases and provide a socially desirable response. For example, the HSI takes a measure of safety through asking the question "Have you ever faced any crises or emergencies in your workplace? How did you respond?" (OPRA group, 2017). The transparency of this item offers the job applicant knowledge on what is being measured and provides them with the opportunity to respond favourably, not necessarily truthfully (Leary & Kowalski, 1990; Schlenker, 1980).

The current psychometric approaches used by organisations to measure workplace health and safety are clearly problematic, being littered with limitations affecting not only the validity of each approach, but also the subsequent selection decisions. Measurement error is extremely important in the safety context. As shown by the high global and national workplace accident statistics, unsafe selection decisions can lead to serious harm. There is a need for a psychometric measure that can measure an applicant's most influential safety behaviours, before commencing employment, without being subject to biases. Gamification of a safety psychometric measure is proposed as a solution to this need.

Gamification

Gamification is "the use of game components in non-game contexts" (Deterding, Sicart, Nacke, O'Hara, & Dixon, 2011, p. 1). From an assessment standpoint, gamification is the use of game dynamics, mechanics, and components in measurements to increase quality and accuracy (Gangadharbatla & Davis, 2016). Development of gamified assessments is new to assessment research. Gamified assessment designs vary greatly, and can look like any standard first-person computer/console game or a computer-based puzzle game.

Gamification and Measurements Bias

Gamified assessments can offer a solution to the biases current safety self-report psychometrics are susceptible to. As outlined above (section: Current Psychometric Safety Measures), for an individual to engage in social desirability or impression management, the following three conditions must be met: measurement context must motivate an individual to present themselves favourably, the construct measured must have a socially desired response, and the construct intended for measurement must be evident to the individual (Furnham, 1986; Villanova & Bernardin, 1991).

The first two criteria will always be met in safety psychometric measures. Given the high-stakes nature of the selection process, job applicants will always be motivated to provide socially appropriate answers and know that being safe is the socially desirable response. However, social desirability and impression management can be prevented through a reduction in item transparency (Furnham, 1986). Through measurement of actual behaviours, gamified psychometric measures reduce item transparency (Aspen Institute, 2007) and can, therefore, prevent the measures' susceptibility to biases.

Measurement of actual behaviours within a gamified environment is a less transparent form of measurement since information is collected through indirect sources. For example, the Hazard Awareness Test *(*HAT) is a gamified psychometric that uses spot-the-difference puzzles to measure hazard awareness (Burt, 2017). Those with greater hazard awareness are more able to find safety-related differences between two pictures supplied (Burt, 2017). This form of measurement is indirect in that spot-the-difference selection decisions do not, on face value, appear to relate to the individual's hazard awareness. Consequently, individuals are less able to intentionally manipulate their answers for a socially favourable response. Through these performance-based indirect measures gamified assessments can gain rich unbiased data, without participants realising how much they are revealing (Reiners & Wood, 2014). Furthermore, gamified assessments have many additional benefits over currently available self-report safety measures which exceed this potentially reduced susceptibility to biases.

Benefits of Gamified Assessments

Gamified assessments can elect authentic behaviours better than traditional paper-based assessments (Clarke, 2009; Clarke-Midura, 2010) and have the potential to provide new insight into the prediction of job performance (Gangadharbatla & Davis, 2016; Shavelson, Baxter, & Pine, 1991). The following three key benefits of gamified psychometrics, over self-report ones, give them their potential for more authentic measurement and greater predictive power: (1) measuring within a more authentic context (Clarke-Midura, 2010), (2) providing more measurable data points than self-report items (Jaffal & Wloka, 2015), and (3) by supporting individual engagement in the selection process (Gangadharbatla & Davis, 2016).

At first glance the incorporation of the gamification elements (reflecting the ability to have multiple attempts supported by a virtual safety net) appear cognitively inconsistent with the concept of authenticity. However, the tasks to be undertaken, the virtual environment, and the measurement points can be designed to be authentic. Research on immersive environments and mediated experiences used in gamification prove that one can create immersive virtual environments, specifically designed, and able to replicate a real-world environment (Ketelhut, Dede, Clarke, Nelson, & Bowman, 2008). Through this virtual replication of real-world environments, game-based assessments are able to cue and capture authentic performance observations (Clarke, 2009; Clarke-Midura & Dede, 2010; Ketelhut *et al*, 2008), perhaps better than current self-report measures (Shavelson, Baxter, & Pine, 1991).

Shavelson and colleagues conducted a series of studies in the 1990s in which they compared computer-simulated performance assessments to paper-based performance assessments (Baxter & Shavelson, 1994; Rosenquist, Shavelson, & Ruiz-Primo, 2000; Shavelson, Baxter, & Pine, 1991). Their findings suggested that hands-on and virtual investigations were tapping into different knowledge than paper-based assessments, and that prior knowledge, behaviours, and experience better influenced how individuals addressed problems within virtual environments (Shavelson, Baxter, & Pine, 1991). Past behaviours are one of the best predictors of future behaviours (Ajzen, 1991). Thus, by electing gameplay decisions that are informed and reflective of applicants' past behaviours, gamification may be a better predictor than self-reports alone.

The second key benefit of gamified assessments over self-report is their multiple points of measurement. Through gameplay logs, game-based assessments are able to collect more measurement items (representations) of one construct than self-report pen-and-paper measures (Gangadharbatla & Davis, 2016). The greater quantity of measurement items provides a higher order understanding of applicant performance/ability on a construct (Darling-Hammond, 2010; Darling-Hammond & Pecheone, 2010; Jaffal & Wloka, 2015). Consider for example Jaffal and Wloka's (2015) gamified assessment '*Bicycle World'*. *Bicycle World* was investigated as a measure of 109 3rd and 4th grade German students' bike safety. To play, participants were given a virtual avatar on a bicycle (observed from the first-person perspective). Through their avatar participants freely navigated throughout the simulated streets while play logs took multiple measures of participants' play decisions, such as: where students choose to make their avatar look while riding, stopping techniques, and traffic rule compliance behaviours.

The play logs provided Jaffal and Wloka with both a detailed and all-encompassing understanding of students' bike safety knowledge. The researchers were able to measure not only students' knowledge of road rules, but their ability to apply these rules in a timely manner and in a variety of real world situations (e.g. different intersection situations). Furthermore, beyond this interpretation and implementation of road rules, the researchers were able to measure students' general bike safety. Given gamification's ability to easily gather multiple points of measurement, play logs also measured presentations of general bike safety behaviours; such as, how long students choose to signal before they turned and how often and in which situations students choose to look behind/around them to check for traffic and other hazards.

In contrast, psychometric measures that use short-answer and multiple-choice items have limited mechanisms for assessing these higher order understandings (Clarke-Midura, 2010; Clarke, 2009). Research has documented that higher-order thinking, skills related to sophisticated cognition (e.g. approaches and responses to novel situations), are difficult to measure with multiple-choice or even constructed-response paper-and-pencil tests (Quellmalz & Haertel, 2004; Resnick & Resnick, 1992; Shavelson, Baxter, & Pine, 1991). By placing individuals in situations that require them to represent their knowledge through their ability to perform in a variety of novel situations, gamified assessments such as *Bicycle World* are able to offer measurement of these higher order safety understandings and provide more detailed and more authentic performance data than what is generally available through self-report items (Herrington Reeves, & Oliver, 2009; Shavelson, Baxter, & Pine, 1991).

The third key benefit is gamification's motivational benefits, which may support greater measurement accuracy (Darejeh & Salim, 2016). Ensuring job applicants are motivated is vital in gathering accurate performance data (Campbell, McCloy, Oppler, & Sager, 1993). According to Campbell, McCloy, Oppler, & Sager's (1993) Job Performance Model, task performance is dependent on declarative knowledge, procedural knowledge, and motivation. Given that selection processes are often perceived as long and tedious, with limited chance of success, applicants can lose motivation. This lowered motivation can result in inaccurate performance scores and the potential for incorrect selection decisions (Campbell *et al*, 1993).

Gamification increases measurement accuracy by developing motivation in job applicants (Campbell *et al*, 1993). Motivational theories, such as Goal-Setting Theory, provide the explanation as to how, via psychometric gamification, applicants' motivation might be improved during selection (Gangadharbatla & Davis, 2016). Goal-Setting Theory describes motivation as the interactive process of reducing the discrepancy between a person's goals and their actual behaviours (Locke, 1968). Goals are at the core of all gamified assessments. Every gamified measure has a task or goal to complete. For example, the HAT described above (in the section Gamifications and Measurement Bias) provides applicants with the goal of finding all the differences between two pictures. Throughout the completion of this task the applicant receives feedback on their performance, such as a click counter and a green box for correct selections. This feedback motivates the applicant by helping them to reduce the discrepancy between their behaviours and their goal (Locke, 1968). Given that motivated applicants are more likely to perform to their best (Campbell *et al*, 1993), gamification's motivational benefits support accurate measurement.

Gamification's ability to better measure authentic performance behaviours in comparison to current paper-pen psychometric measures presents clear benefits for its use in psychometric measurement. Given these benefits of gamified assessments over self-report, a gamified psychometric is proposed as a solution to the limitations of currently available self-report safety psychometric measures.

Safety Behaviour Test (SBT)

The SBT, a gamified safety behaviour test, which is developed and examined in this study is proposed as the solution to current health and safety measures. Current safety psychometric measures' susceptibility to common biases, measurement of less influential factors in accident causality, and less predictive power than performance-based measures, make their use for selection questionable. The SBT's measurement of safety behaviours capitalises on the gamified measure's predictive power by measuring the most influential factors in workplace accidents (see Figure 1) (Ramsey, 1985; Reason, 1990; Christian *et al*, 2009), while potentially reducing the test taker's susceptibility to social desirability and impression management biases.

The SBT is fully described in the method section, but briefly the SBT is a fully animated computer game of the point-and-click genre. Within the SBT, individuals are tasked with navigating a forklift through a warehouse to collect and transport various objects from their stored locations to a waiting truck for transportation. Throughout the course of the game each player encounters various situations, which provide opportunities to display (or not) safety behaviours. The potential benefits of SBT's bias-reducing capabilities and more accurate measurement, however, are only relevant to the extent the gamified measure is valid. Hence the key question addressed in this research.

Gamification Validity

To be of value the SBT must be valid, meaning it must accurately measure safety behaviour construct(s) and be predictive of future safety performance (American Psychological Association *et al*, 1999). Very few studies have investigated the implementation of gamified assessments, and even fewer have investigated the validity of gamified measures. Of those which do report on gamified assessments validity, results are mixed. Some studies have reported their gamified measures were not valid (Jaffal & Wloka, 2015; Whetzel, McDaniel, & Pollack, 2012), while others report they were able to develop a valid gamified measure (Burt, 2017; Mislevy, Almond, & Lukas, 2003; Kim & Shute, 2015), supporting the use of a gamified measure as a way to observe authentic behaviours (Clarke, 2009; Clarke-Midura & Dede, 2010; Ketelhut *et al*, 2008; Shavelson *et al*, 1991). These mixed results may not necessarily reflect on assessment gamification as a whole, but may be indicative of the design of the measurement points within the specific gamified assessment(s) investigated.

To understand the potential for variation in validation results, it is necessary to understand that gamified assessments contain two key design components, measurement design and gamification design. Measurement design is developed first and contains the construct intended for measurement and the design of its measurement items. During the gamification design phase, the virtual environment is created and the measurement items are applied to decision points within the game (Halverson & Owen, 2014; Ifenthaler, Eseryel, &

Ge, 2012). Each decision point is recorded by a play log and summed with the other associated decision points to provide the user with a score on the measured construct.

Game-based assessments have a large variability in terms of constructs intended for measurement and the design of the items through which these constructs are captured or measured. To clarify with an example, the HAT is a gamified measure of hazard awareness using gamified features such as spot-the-difference items of leaking gas and broken lights to measure employee hazard awareness (Burt, 2017) (as noted above in section: Gamification and Measurement Bias). The HAT, however, will not be guaranteed to be valid just because gamification has been applied, the items of measurement must be accurate (Burt, 2017). The validity of a gamified assessment may vary depending on the sophistication of the gamification design and/or the features of the assessment which have been designed to measure the construct of interest (Kim & Shute, 2015; Werbach & Hunter, 2012).

In part, a game-based assessment's validity and ability to predict employee behaviours is centred on its similarity to a work-sample (Gangadharbatla & Davis, 2016). Work-samples have been consistently regarded as one of *the most* accurate and valid measures in predicting performance (Hunter & Hunter, 1984; Reilly & Warech, 1993; Roth, Bobko, & McFarland, 2005) over interviews, application blanks, and self-report psychometrics (Schmidt & Hunter, 1998). Work-samples, commonly used in selection, require an applicant to provide a sample of their ability to perform on a construct (task) – which is then scored.

Of course, it is difficult, if not impossible, to use work-sample testing to measure safety constructs. For example, to measure the constructs of safety compliance and safety participation, a work-sample would require an individual to complete a potentially dangerous task, such as using a chainsaw, while being watched by another individual who would record them on their actions of safety compliance (using earmuffs) and participation (providing earmuffs to those around them). It would be unethical and dangerous for a selection process to require employees to engage in these physical work-samples in order to obtain information on their safety behaviours.

In contrast, a gamified assessment can replicate a work-sample by requiring the individual to provide a performance-based sample through a virtual, rather than physical, environment. Through the game lens of the SBT an applicant is provided with a medium to provide a sample of their safety performance behaviours in a safe virtual environment. However, before the SBT can be used as a psychometric tool it must first be established as a criterion-related valid gamified measure.

Criterion-Related Validity

A selection tool that has criterion-related validity is one that is effective in hiring people who can perform the job to a certain standard. Criterion-related validity pertains to evidence of a "relationship between the attributes in a measurement tool with its performance on some other variable (the criteria)" (DeVon, 2007, p. 1558). Standards for Educational and Psychological Testing, developed jointly by the American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, outline criterion-related validity as a foundation to any psychometric measure. If a psychometric tool does not accurately predict the construct it claims to, the tool is inadequate - and unable to serve as a psychometric measure (American Psychological Association *et al*, 1999).

The classic approaches to the establishment of criterion-related validity is a concurrent or predictive method of validation (Cascio & Aguinis, 2005). The predictive validity approach involves a time interval where the measure's validity will become available at a later time; for example, a safety measurement tool is used and then at a later date performance criteria (i.e. safety behaviours) are measured. The concurrent validity approach presents only the status quo at one time. For example, a safety measurement tool is used and then an already validated measure of the same safety behaviours is used as the criteria. Generally speaking, both designs involve the same paradigm in which a relationship (correlation) is established between predictor scores and data on the criteria (standard already established as an accurate representation of the construct in question) (Cascio & Aguinis, 2005). The strength of the correlation substantially supports the extent to which the instrument measures performance on each criterion (Waltz, Strickland, & Lenz, 2010). For example, the criterion-related validity evidence of work-samples is reported to have the highest relationship with performance behaviours of any psychometric tool of up to $r = .54$ (Schmidt & Hunter, 1998).

Although the current study is more reflective of a concurrent validation approach, neither a concurrent nor predictive validation approach is strictly followed. The main difference is that the current study’s sample is not taken from one organisation or one job. However, the general approach of both the predictive and concurrent approach is followed to obtain validity evidence of the SBT, in which the relationships is examined between the predictor measure (the SBT) and the criterion measure(s) (Cascio & Aguinis, 2005). The criteria measures used to validate the SBT were safety behaviour measures completed by an independent source who had knowledge of the test taker.

Current Study

The current study aims to establish criterion validation for the SBT, so that it may (in the future) be used as a psychometric tool in replacement of current biased and less accident predictive measures. To achieve this aim the following research question was investigated: Does the SBT have criterion-related validity? SBT criterion-related validity evidence was established through comparison of SBT participant scores on the SBT and their corresponding acquaintance's independent safety behaviour scores from already validated research measures. Note that these scales were developed to measure within a research low-stakes environment and they are not validated for use in a selection/psychometric context. In the investigation of this research question opportunities were available to examine the construct validity of the SBT and the relationship between acquaintance informed safety behaviour criteria and accident outcomes. No specific hypotheses were formed for the study, however, the results were expected to support the SBT having evidence of criterion-related validity.

Method

Design

This study used a concurrent criterion-related validation design, with the exception that sampling was not conducted within one organisation. In order to gather an appropriate sample size participants were recruited from a variety of employee groups. The SBT was administered followed by two questionnaires: the *individual characteristics questionnaire* (ICQ) (the full ICQ is shown in appendix B) completed by the individual that took the SBT, and the *acquaintance questionnaire* (AQ) (the full AQ is shown in appendix A) completed by an acquaintance nominated by the individual that took the SBT. Note, the questionnaires in appendix A and B identify the measurement tool under investigation as the Compliance and Participation Test (CPT), not the SBT. This was because the name of the test was changed during this study.

To investigate the research question "Does the SBT have criterion-related validity?", SBT scores were correlated with the data on safety behaviours obtained from the AQ. Data obtained from the ICQ was primarily used by Thomas (2018) who investigated whether a person's characteristics show any *adverse impact* on SBT performance. However, the measures of safety compliance, safety participation, safety voicing, safety consciousness, risk-taking, and rule bending obtained in the ICQ were used by the current study to demonstrate that self-report data on safety behaviour can show bias in the manner expected. This dissertation also accounted for the findings from Thomas (2018), in that variables identified as showing adverse impacts on SBT performance were controlled for when examining SBT criterion-related validity. The current study was reviewed and approved by the University of Canterbury Human Ethics Committee, Reference number HEC 2017/26.

Participants

Recruitment of SBT Participants

In order to obtain SBT participants, haphazard sampling was used, in which the most available people were studied (Weisberg & Bowen, 1977). The only recruitment criteria was that SBT participants were in full-time or part-time work and had to agree that they had adequate vision to see a computer screen. As a result of the haphazard sampling, participants were recruited from New Zealand organisations and the University of Canterbury student population.

SBT participants obtained from organisations were recruited through email and phone. These communications contained information regarding the study’s purpose, process, ethical approval, and what the organisations could gain from allowing their staff to participate (see appendix C). An advertisement was used to recruit participants from the University of Canterbury student population. This included a brief description outlining the study’s purpose, who could participate, and the reward offered for participation (see appendix D). All SBT participants received a $10 MTA petrol voucher upon completion of their participation.

Recruitment of Acquaintance Participants

Acquaintances were recruited by the SBT participant. The SBT participants were told when selecting their acquaintance, to participate, the individual must have knowledge of their safety behaviour. Acquaintance participants were either a supervisor, friend, family member, or co-worker of their respective SBT participant. Before agreeing to participate the recruited acquaintance was provided with an information sheet (see Appendix G) outlining the purpose and process required for their participation in the study. Acquaintances also received a $10 MTA petrol voucher for their participation.

Participants

The study consisted of 200 participants in total, 100 of which participated as SBT participants, while the remaining 100 participated as acquaintances/independent data sources. 9% of the 200 recruited participants were recruited from the university student population, while the remaining 91% were recruited from New Zealand organisations. Table 2 displays the descriptive statistics of the participants. As evident in Table 2 one acquaintance participant chose not to provide information on their gender.

Table 2

Demographic Information of Participants by Participation Group

Variables	SBT Participants *n*= 100	Acquaintance participants *n*=100
Males	62	50
Females	38	49
Mean age (*SD*)	41.64 (14.19)	43.56 (14.60)
Age range	18-66	18-66

Materials

Materials for this study were split between SBT participants and their acquaintances. Materials used for SBT participants were the SBT computer game and the ICQ, which included general demographic questions and measures of the following safety behaviours:

safety compliance, safety participation, safety voicing, safety consciousness, risk-taking, and rule breaking (see appendix B). The sole material used for acquaintance participants was the AQ (see appendix A). The AQ included general demographic and accident quantity questions, in addition to questions regarding the nature of acquaintance's relationship with their respective SBT participant, followed by the same safety behaviour measures used in the ICQ to measure the SBT participant's safety behaviours. Measurement titles/description of the safety behaviour measures was not included in the questionnaires when they were used for data collection. However, for the clarity of the reader, in Appendix A and B measurement items have been labelled with the type of safety behaviour they measure.

SBT Participant Material: Safety Behaviour Test (SBT)

The SBT uses a gamification design to measure safety behaviours. The SBT is a fully animated computer test of the point-and-click game genre, meaning that the test taker can point the cursor at an area on the screen and click in order to interact with the test environment. First person views are used in the SBT, thus the test taker experiences the test as if they were a character navigating the test environment. To protect the security of the SBT a full version of it is not provided as part of this dissertation. In addition to which a full description of the test and the specific measurement points are not given, as the confidentiality of the test needs to be maintained. Should a description of the test become publicly available, the usefulness of the test becomes significantly reduced (Burke, 2009). As this dissertation will become publically available within the University of Canterbury Library, writing a description of the SBT and the measurement points within it will threaten the security of the test. What follows is a description of the main features of the SBT.

The SBT requires the test taker to assume the role of a worker in a waste disposal company. In the SBT, players are given instructions to retrieve several different items from a warehouse with a forklift, and then load those items into a container. While a number of different tasks would have provided a medium in which to measure safety behaviours, the loading scenario was chosen as it was thought that using a forklift is an activity that most people can understand in a gamified setting.

A number of game design elements were included in the SBT. A timer was shown in the top left corner of the screen displaying the play time in seconds along with a click counter displaying the number of times the player had clicked the mouse, red crosses appeared after an incorrect choice is made within the game, all of which provided feedback to the test taker (shown in Figure 3).

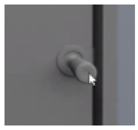
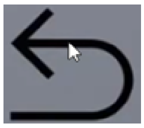
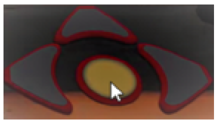
The SBT is a standalone program that is hosted on the cloud and uploaded through Chrome Web Browser. A Lenovo ideapad 510-15ikb laptop was used to run the SBT for the duration of the study, which had a 15.6 inch screen. The test began with the instruction page to explain how to interact with the game environment (see Figure 2). The instruction page showed images of the test, accompanied by explanations of how to navigate those areas. For example, a door handle was shown with a mouse cursor on top of it, with a description that explained that the participant needed to point-and-click on the door handle to make the door open. At the bottom of the instruction page was a button that read “START”. The instructions explained that once the participants had finished reading the instructions, they needed to click a button to begin the test.

Figure 2

Instruction Page for the SBT.

Test Instructions

Before you begin the test it is important that you understand how it works. Please carefully read the following points.

• This test is a work simulation in the format of a point and click video game. In the game you play the role of forklift driver number 1. You will enter a building which contains items you may interact with by clicking on them with the mouse pointer. For example to open a door, click on the door handle.	
• It is important to note that it may not be possible to go back in the game after clicking certain things such as a door handle, as this action will move you to the next area within the game. However, in some sections of the game a back arrow will appear in the bottom left corner of the screen. Clicking this will move you back in the game.	
• For a large part of the game you will be in a forklift. When you are in the forklift you can only control the game by clicking areas on the forklift control panel. **For example to select an item location, click on the level location on the control panel.** Controls that you can use within the forklift will change at various parts of the game.	
• The forklift directional control arrows will always be present. Click these in the middle of the arrow to control directional movements. You can only control directional movements when the forklift is stopped.	
• At one point in the game a yellow stop button will appear on the control panel. This button allows you to stop the forklift.	
• You can only control the test (e.g., select directional movements of the forklift) when the test is in **manual** mode. Clicking anything when the test is in **auto** mode will have no effect, and only waste clicks. Test mode is shown in the bottom middle of the screen.	**AUTO** **MANUAL**
• A mouse click counter is shown in the top left corner of the screen. **The test can be completed <u>perfectly</u> in 50 mouse clicks.**	Clicks Made 17/50

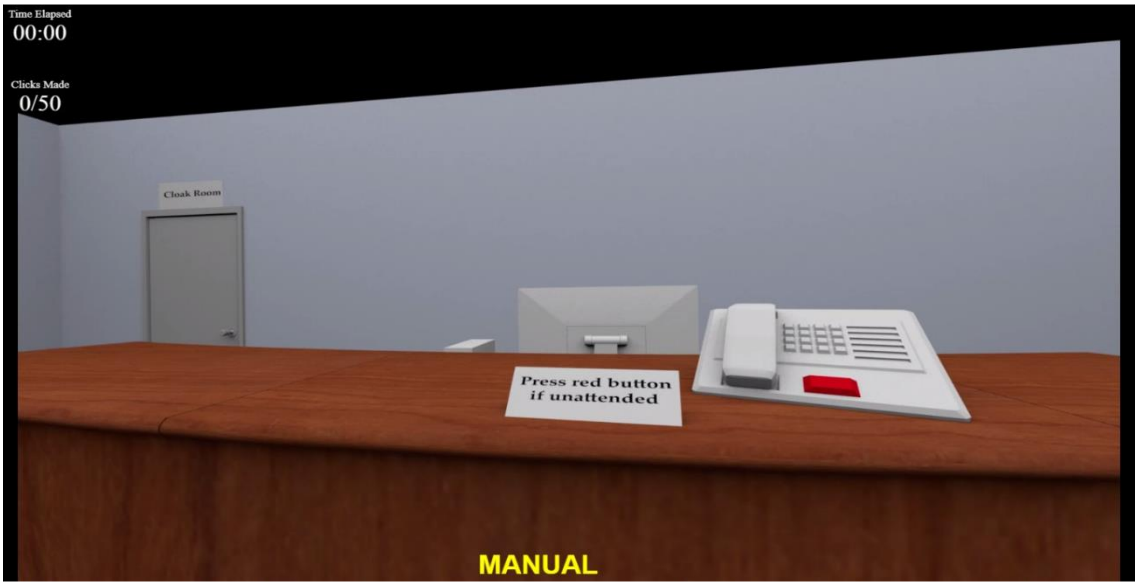
- Further instructions on what you need to do will be given when the game begins.
- Please close other tabs and don't have applications running in the background.
- When you click start the test will take several minutes to load. After which it will automatically start.

START

After clicking the "START" button on the bottom of the instruction page, the participants were presented with an office scene (see Figure 3). A sign on the desk in the scene read “Press red button if unattended”. After pressing the red button on the phone, the participant heard the following audio: “*Hello forklift driver number 1. Sorry, I am up on level 6. It’s good that you are here on time, there is only one job for you today. You have a shipment for disposal at the incinerator. The empty shipping container for the shipment is in loading dock C. A truck will take the loaded container to the incinerator as soon as you have finished loading it. I have already put the shipment items into the system, so when you get in a forklift the item list will be on the display screen. The new semi-automatic forklifts are working great, just click an item on the list and off you go to the relevant floor. Remember that control buttons appear when you need them. We have fixed the problem with the red right and left directional control arrows, and the central yellow stop button is working fine on all forklifts. Remember to load the items in the order shown on the list. The cloakroom is nice and tidy this week, so let’s keep it that way. Don’t muck around as the transportation firm will charge us if they have to wait, but be careful. When you have got the order loaded come back here and let me know. If you would like me to repeat the instructions, just click the red button again*”.

Figure 3

Showing the Office, The Opening Scene of the SBT.



The word “MANUAL” or “AUTO” appeared at the bottom middle of the screen. When the word ‘manual’ is shown the player can click an area in the game to make a decision. When the word ‘auto’ is displayed the player has no control over the game – it is essentially in auto-play mode. During the SBT the player made 35 play decisions that were recorded and saved in a data file during test completion. Although specific information on the 35 measurement points is not provided here in order to maintain test security, these measurement points were designed to reflect behaviours falling under the following safety behaviour categories: safety compliance, participation, voicing, and consciousness. During the game development phase logical combinations of these measurements were examined to determine if they formed measurement scores of safety behaviours. As you can see in Table 5, 13 decision points were combined to provide an SBT score.

SBT Development and Piloting

Over a three week period the SBT was piloted with a range of different people in order to assess test usability. A haphazard sampling method was used to obtain 10 participants for the pilot study. In order to measure the usability of the test, participants completed the test in view of the researchers and were asked to discuss how usable they felt the test was once they had finished. The 10 participants that took part in the pilot study were each given a Warehouse voucher to the value of $10 for their time. As a result of the pilot study the researcher undertook usability development on the SBT. The written instruction page shown in Figure 2 was created to increase clarity. Furthermore, slight alterations were made to the test itself, such as allowing the verbal instructions, played after clicking the red button on the phone, to be heard for a second time and the auto/manual sign shown in Figure 3 was added.

In order to determine if the measurement points of the test were being accurately recorded and saved onto the data file, the researcher completed the test using a range of different play decisions and manually recorded which specific decisions were being made correctly. The manual scoring sheet was compared to the data file produced by the SBT, and changes were made to the scoring system in the test based on any discrepancies found. This process of SBT development was continued until the SBT was deemed to be correctly recording measurement points. Subsequent to these developments the SBT was considered ready for criterion-related validity investigation.

Acquaintance Material: Acquaintance Questionnaire

Biographical and Relationship Data

Biodata information requested the acquaintance's age and gender. The type of relationship the acquaintance has with their respective SBT participant was also measured, including how long the acquaintance has known them for (months) and how well they knew them on a scale anchored at 'Not very well at all' = 0 to 'Extremely well' = 100. A high score was indicative of a greater knowledge of their respective SBT participant. Questions and response formatting are provided in (Appendix A).

Rated Safety Risk Score

A 100-point scale measured SBT participants' general degree of safety risk. The scale was anchored at 'Not at all risky' = 0 and 'Extremely risky' = 100, where a high score was indicative of a greater degree of risky behaviours in all situations. Questions and response formatting are provided in (Appendix A).

Accident Frequency

Three items measured the frequency of accidents and incidents over the SBT participant's entire life. These items required the acquaintance to indicate the frequency of near-miss accidents (which could have resulted in injury or damage), minor injuries (requiring medical attention), and lost time injuries (which required time off work) that the SBT participant had experienced within three different locations (work, home, and other). A fourth column was added, labelled “don’t know”, and this column was used should acquaintances not have accident frequency knowledge. Data from this section of the questionnaire was not used in the analysis reported in this dissertation. The specific questions and response formatting are shown in (Appendix A).

Safety Behaviour Scales

Safety behaviour was measured using six scales frequently used in safety research: safety compliance, safety participation, safety voicing, safety consciousness, risk taking, and rule-bending. The wording of items were adjusted from a first-person format to suit third-person acquaintance reports on an individual. For example, “*I always use all the necessary safety equipment to do my job”* was adjusted to *“*...always uses.... ”.* Acquaintances were instructed that the * referred to the individual that asked them to complete the questionnaire. Each scale item was responded to on a 5-point Likert scale anchored at 'strongly disagree’ = 1 and ‘strongly agree’ = 5. Scales were analysed for internal consistency and coefficient alphas reported. Only one scale indicated the need to remove an item to improve reliability. Details of this analysis and examples of scale items are given below. Scale scores were formed by summing each safety behaviour’s item rating and dividing the sum by the number of items in the scale. Depending on the scale, a higher score is indicative of higher levels of safety, or a higher level of risk-taking behaviour.

Participation and Compliance

Safety compliance and participation were measured using an adapted version of Neal and Griffin’s (2006) six-item scale. Three of these items measured safety compliance, core activities an employee needs to engage in to maintain workplace safety (Neal & Griffin, 2006). The other three items measured safety participation, behaviours which help to develop an environment that supports safety (Neal & Griffin, 2006). An example item for safety participation is “*At work *…puts in extra effort to improve the safety of the workplace”* and an example of a safety compliance is *"*...always uses all the necessary safety equipment to do their job".* Participation and compliance items were summed separately to give each participant two scores, each ranging from 3-15. The two subscales, compliance and participation, were reported by Burt, Banks, and Williams (2014) to have excellent coefficient alphas - with the respective coefficients of .93 and .86. The current study reported coefficient alphas for safety compliance and participation of .87 and .83, respectively.

Safety Voicing

Acquaintances' perception of their respective SBT participant's safety voicing behaviours was measured using an adapted version of Tucker *et al*'s (2008) five-item Safety Voicing Scale. Safety voicing is defined as "any individual communication directed at improving safety conditions" (Tucker *et al*, 2008, P.319). An example item is "*... makes suggestions about how safety could be improved*". Safety voicing scores had a possible range of 5-25. The original scale by Tucker *et al* (2008) reported an acceptable coefficient alpha of .78. The current study reported a safety voicing coefficient of .87.

Safety Consciousness and Risk-taking

Safety consciousness and risk-taking were measured using the 12-item Safety Consciousness and Risk-taking scale developed by Westaby and Lee (2003). Where safety consciousness is defined as "a positive attitude and awareness toward acting safely in general", and risk-taking is an "individual's willingness to engage in activities that knowingly have elements of physical danger" (Westaby & Lee, 2003, p. 228). An example of safety consciousness is *"*... gets upset when seeing other people acting dangerously*"; and an example of risk-taking is "*... values having fun more than being safe*". Scores ranged from 7-35 for safety consciousness, while risk-taking ranged from 5-25. Westaby and Lee (2003) reported a good coefficient alpha of .85 for the safety consciousness items and an acceptable coefficient alpha of .77 for the risk-taking items. The current study reported an acceptable safety consciousness coefficient of .82 and in initial risk-taking coefficient of .79. Removal of one item from the risk-taking scale increased the coefficient to .82.

Bending the Rules

The propensity for the SBT participants to breach workplace safety rules and procedures were measured using an adapted version of Chmiel's (2005) four-item Bending the Rules Scale. Where rule bending was defined as the tendency to break rules and breach safety procedures. This was measured through items such as *"*... sometimes cuts corners if it makes the task easier"*, with scores ranging from 4-20. The original scale by Chmiel (2005) reported a good coefficient alpha of .82, the current study reported a coefficient of .87.

SBT Participant Material: Individual Characteristics Questionnaire

A full description of the ICQ is given in Thomas (2018). However, those ICQ variables which are included in the analysis reported in this dissertation are described below.

Computer Game Experience

Four items measured SBT participants' computer game experience. These items required the SBT participant to initially indicate whether or not they had played computer games before and, if yes, provide numerical information on how many years and months they had played them for, along with how frequently they played. For this measure, a higher score showed that the participant has had more computer game experience. The specific questions and response formatting are shown in (Appendix B).

Accident Frequency Rates

Three items measured frequency of accidents and incidents. These items required the SBT participant to indicate the frequency of near-miss accidents (which could have resulted in injury or damage), minor injuries (requiring medical attention), and lost time injuries (which required time off work) that they have experienced within three different locations (work, home, and other). These frequency ratings were a numerical value and ranged from 0 upward. For analysis, the responses for each category are summed over the three locations. High scores in each category suggested that the participant has had more safety-related near-misses, accidents, or incidents. Questions and response formatting are provided in (Appendix B).

Health and Safety Training

Three items measured SBT participants' health and safety training experience. These items required SBT participants to report on whether they had completed Health and Safety training, and if so how many different types of health and safety training programmes they had completed, along with a numerical measure of how many hours these training programmes combined to. For this scale, a higher score will suggest that the participant has had more health and safety related accident training. Questions and formatting are provided in (Appendix B).

Safety Behaviour

The same four scales were used to measure six safety behaviours in the ICQ, as in the AQ. These were the following scales: Neal and Griffin's (2006) Safety Compliance and Participation Scale, Tucker *et al*'s (2008) Safety Voicing Scale, Westaby and Lee's (2003) Safety Consciousness and Risk-taking, and Chmiel's (2005) Bending the Rules Scale. Items in the ICQ were worded in the first-person. For example, "*I always use all the necessary safety equipment to do my job*". All other aspects of the scales, including response formatting and scale scoring, were identical to that used in the AQ. Thomas (2018) reported scale coefficient alpha results for the ICQ safety behaviour measures.

Testing Procedure

The collection of SBT data occurred under the supervision of the researcher. University of Canterbury participants were met outside Psychology Room 607 as per the mutually agreed time. Participants recruited as part of an organisation were tested at their respective workplace.

All SBT participants were presented with an information sheet (see Appendix E) and a consent form (see Appendix F). Following their signed consent, participants were allocated a code that was used to record their SBT, ICQ, and AQ data under. Then participations were guided into an office where the SBT was opened on a laptop to the instruction page, ready to be used. It was paramount that this environment was practical for laptop use and limited in noise exposure. Before beginning the test SBT participants were told the following to ensure they all attempted to perform their best on the test: *"This is the instruction page of the test. Please read the instructions carefully, as you will only be able to see them once. Press the start button when you are ready to take the test. I will leave you to take the test privately, and will be waiting outside of the room for when you have finished. Please imagine that you have applied for a job. The test you are about to complete is being used to determine your suitability for the job. As a job applicant, try to do your best on the test".*

After completing the SBT, SBT participants came to collect the researcher and were presented with the ICQ, which they then completed within the same office previously used for the completion of the SBT. After this questionnaire was completed the SBT participant received an unsealed envelope containing an acquaintance information sheet (see Appendix G), a consent form (see Appendix H), and the AQ (see Appendix A). The SBT participant then gave this envelope to their selected acquaintance, who completed the forms, sealed them back in the envelope, and then returned to the researcher. Once both parts of the study were completed the SBT participant and their acquaintance were given a $10 MTA voucher for their participation.

Results and Discussion

Data Management

SBT data was downloaded into a Microsoft Excel file and subsequently entered into an SPSS file to complete the analysis. The data recorded from the ICQ and the AQ were combined with the SBT information in SPSS. Data inspection found some outliers and cases of missing values within the AQ and ICQ. The missing data was replaced with the item mean when less than 10% of the data was missing and the missing data was not a biographical response. Table 3 provides information on the quantity of missing values and the replacement mean, where a replacement mean was used.

Table 3

Missing Items and Replacement Means for the Acquaintance and Individual Characteristics Questionnaire

Measurement items with missing responses	Missing item	Number of missing items	Item mean
Acquaintance Questionnaire			
Length of time known		1	
Safety compliance	Item 1	1	4.15
	Item 2	1	4.15
Individual Characteristics Questionnaire			
Play computer games		2	
Months playing computer game		3	
Safety compliance scale	Item 3	1	4.25
Safety participation scale	Item 2	2	3.85
Safety voicing scale	Item 1	6	3.75
	Item 2	1	4.10
Safety consciousness scale	Item 1	1	3.83
	Item 3	2	3.83
Rule breaking scale	Item 1	3	2.55
	Item 4	2	2.29

Given the ability of extreme responses to distort results, outliers were removed from all data sources, using the rule of more than plus or minus three standard deviations away from the mean (Field, 2009). A total of two responses were removed from two different measurement items, as shown in Table 4. Descriptive statistics for the AQ, ICQ, and SBT are presented throughout the results as their corresponding data was analysed.

Table 4

Individual Characteristics Questionnaire Outliers

Individual Characteristics Questionnaire Outliers	Number of outliers removed	*M* (*SD)* After outlier removal
Number of different health & safety training	1	3.41 (2.26)
Hours of health and safety training	1	16.90 (22.04)

Distribution and Range Restrictions in the SBT score

Before the SBT's criterion-related validity evidence was examined the *SBT score* data was first investigated. Table 5 shows the descriptive statistics for the *SBT score*. Specifically, Table 5's analysis focused on the current studies ability to report accurate SBT construct and criterion-related validity evidence through investigation of the *SBT score*'s distribution, range, skew, and kurtosis. Examination of Table 5 reveals no range restrictions and relatively normal skew and kurtosis, which show the *SBT score* results do not deviate from the normal. These findings are desirable, given the current study investigated construct and criterion-related validity evidence through correlational analyses that require normally distributed data (Hunter, Schmidt, & Le, 2006).

Table 6 displays SBT participants' distribution in performance on the *SBT score* measure. Expected frequency results for a measurement tool are for scores to be normally distributed, and thus have a large grouping of scores in the middle with fewer scores distributed at the top or lower end of the scale. Inspection of Table 6 reveals that the *SBT score* distribution results were reflective of these expected characteristics. Scores were distributed throughout the entire possible range and the majority of SBT participants performed in the 'middle range', with few scores falling within the top or low range. These results are also reflected in Table 5's percentile score data, whereby only a quarter of participants scored at either extreme of the measurement scale.

Table 5

Descriptive Statistics for the SBT Score Measure

SBT measure	*n*	*M* (*SD*)	Range: Possible	Range: Actual	Skew	Kurtosis	Percentiles SBT score (% below)
SBT score	100	8.05 (8.00)	0-13	0-13	-.45	-.06	6 (25%) 8 (50%) 10 (75%)

Table 6

Frequency of SBT Participant's Performance on the SBT Score Measurement

SBT score	Frequency and percentage of scores *n=100*
0	1
1	1
2	1
3	1
4	8
5	4
6	12
7	15
8	8
9	15
10	14
11	12
12	5
13	3

Distribution and Range Restriction in the Criterion Variables

As with the SBT, the safety behaviour criterion data was examined for any distribution or range restrictions. The criterion was obtained from acquaintance reports on SBT participants' safety behaviour using the following six scales: safety compliance, safety participation, safety voicing, safety consciousness, risk-taking, and rule bending. Table 7 presents descriptive statistics for the AQ safety behaviour measures. Inspection of the table shows no range restrictions, with acquaintances reporting to both ends of the scale. Results did, however, reveal some large negative skews and large positive kurtosis for the criterion measures of safety voicing and safety compliance. The large skew and kurtosis showed that the data from these measures deviate from a normal distribution. This finding is undesirable given correlations will be used to establish SBT criterion-related validity evidence, which, as outlined above, require normally distributed data. The deviations of these measures from a normal distribution will likely suppress any correlations found between these criterion variables and the *SBT score* when establishing criterion-related validity evidence (Hunter, Schmidt, & Le, 2006).

Table 7

Safety Behaviour Measure Descriptive Statistics the Acquaintance Questionnaire (AQ)

Safety behaviours measures	*n*	Range (min – max)	Skew	Kurtosis
Compliance	100	0 - 5	-1.33	2.39
Participation	100	0 - 5	-.92	.63
Voicing	100	0 - 5	-1.27	1.88
Safety consciousness	100	1.14 - 5	-.59	.14
Risk-taking	100	0 - 4.75	.82	.90
Rule bending	100	0 - 4.50	.083	-.80

Criterion Variable Relationships

As outlined in the introduction (section: Models of Safety), safety behaviours of voicing, compliance, participation, and consciousness collectively describe safety behaviour. As such, these measures were specifically selected to examine the criterion-related validity of the SBT. Table 8 shows correlations between the criterion measures. Inspection of Table 8 reveals significant and positive correlations between the measures of safety behaviours and significant negative correlations between safe and unsafe behaviour measures. However, the inter-correlations are not so great as to suggest one or more of the safety behaviour scales are redundant.

Table 8

Correlations Between Criterion Safety Measures

Criterion measures	*1*	*2*	*3*	*4*	*5*
1. Compliance	-				
2. Participation	. 44**	-			
3. Voicing	.39**	.74**	-		
4. Safety consciousness	.53**	.66**	.66**	-	
5. Risk-taking	-.16	-.10	-.05	-.31**	-
6. Rule bending	-.37**	-.14	-.20*	-.33**	.57**

*p<.05. **p<.01.

Distortion in Self-report Data

Before examining for criterion-related validity, a key assumption behind the development of the SBT was examined. The assumption which prompted the development of the SBT was that individuals will distort their self-reports on safety behaviour measures based on social desirability and impression management processes. The distortion was predicted to result in the inflation of positive safety behaviours and the reduction of negative safety behaviours. To investigate this, a T-test and correlation comparison was run between the ICQ and AQ safety behaviour ratings. Table 9 presents the findings of these analyses.

Inspection of Table 9 reveals relatively small significant, but relatively small correlations between the two different data sources. These significant correlations are what one would expect if there is systematic distortion or bias in the SBT participants' data. Examination of Table 9's T-test results shows significant differences between the SBT participant and acquaintance data for three of the safety behaviour scales and differences approaching significance for two other scales. Examination of the significant means indicates an overestimation of safety behaviours on the part of the SBT participation in comparison to the acquaintance. Together, the T-test and the correlation results support the proposition that individual self-ratings on safety behaviours are likely to contain a degree of error. This is consistent with the notion that self-report measures are influenced by biases aimed towards presenting oneself favourably.

Table 9

T-test and Pearson Correlational Comparison between the Individual Characteristics Questionnaire (ICQ) and the Acquaintance Questionnaires (AQ) Measures of Safety Behaviours

Safety behaviours measures	ICQ Self-report mean (*SD*)	AQ Independent report mean (*SD*) *n=100*	*t* (99)	*p*	Pearson correlation between ICQ and AQ
Compliance	4.25 (.67)	4.03 (.95)	-2.11	.03*	.26*
Participation	3.87 (.81)	3.52 (1.20)	-2.84	.00**	.28*
Voicing	3.90 (.64)	3.67 (1.12)	-2.10	.03*	.31**
Safety consciousness	3.79 (.60)	3.63 (.78)	-1.76	.08	.22*
Risk-taking	2.24 (.95)	2.06 (.09)	-1.78	.07	.45**
Rule bending	2.52 (.91)	2.35 (1.00)	-1.29	.19	.14

*p<.05. **p<.01.

Criterion Measure Consistency

Table 11 shows the results of correlating acquaintances' safety behaviour criterion measures with their rated safety risk score (i.e. their score on the following question "Considering *…'s behaviour in all the situations that you know, please indicate *…'s general degree of safety risk by placing a mark on the 100 point scale." 'Not Risky at all' = 0 or 'Extremely Risky' = 100 Risky). As shown in Table 10, the SBT participant sample had diverse safety behaviours, reaching both ends of the 100-point scale. If acquaintances could report on their SBT participant's safety behaviours accurately, their data should be consistent; scores on the 100-point rated safety risk scale should correlate with data on the specific safety behaviour measures. Inspection of Table 11 reveals consistency in acquaintance reports. SBT participants' rated safety risk scores correlated negatively with measures of safe behaviours (i.e. the higher rated the safety compliance score, the lower the safety risk score) and positively with measures of unsafe behaviours. This consistency is indicative of a degree of accuracy in the information provided by acquaintances.

Table 10

Descriptive Statistics for the Acquaintance Questionnaire Rated Safety Risk Data

Measure	*n*	Minimum	Maximum	*M* (*SD*)
Safety risk score (100-point scale)	100	0.00	100.00	32.32 (30.13)

Table 11

Correlation of the Acquaintance Samples' Rated Safety Risk 100-point Scale and Other Safety Behaviour Measures

Safety behaviour measures	Rated safety risk
Safety compliance	-.44**
Safety participation	-.36**
Safety voicing	-.30**
Safety consciousness	-.47**
Risk-taking	.22*
Rule bending	.36**

*p<.05. **p<.01.

SBT Participant and Acquaintance Relationship

Before undertaking the validation analysis, the relationship between the SBT participant and their acquaintance was examined. The quality of the independent criterion data is likely to be highly dependent on the strength of the relationship, as to rate a person's safety behaviour, the acquaintance has to know them reasonably well. Table 12 and Table 13 provide descriptive information on the type of relationships reported between SBT participants and their respective acquaintances. Inspection of Table 12 shows that the majority of acquaintances knew their SBT participant through a workplace or partner/spousal relationship.

Table 13 reveals these relationships as close, with both measures (*length of time known* and *how well known)* reporting high averages with maximums of the *how well known* measure reaching the top of the 100-point scale. Together, Table 12 and 13's findings support the usefulness of acquaintance participants as informed sources.

Table 12

Descriptive Statistics Outlining the Type of Relationship between Acquaintances and their Respective SBT Participant

Acquaintances relationship type	Frequency
Work colleague	47
Work manager or supervisor	13
Spouse or partner	22
Child	3
Parent	2
Friend	13

Table 13

Descriptive Statistics of the Relationship between Acquaintances and their Respective SBT Participant

Variables	*n*	*M* (*SD*)	Minimum	Maximum
Months known	99	144.79 (172.90)	1	780
How well known (100-point scale)	100	73.46 (22.15)	10	100

SBT Construct Validity

A test cannot have criterion-related validity if it does not display any evidence of construct validity. If the SBT has construct validity as a measure of safety behaviour, then the *SBT score* should be influenced in a predictable way by variables associated with safety behaviour. ICQ data provided an opportunity to examine the construct validity of the SBT by correlating the participants *SBT score* with reported health and safety training. Furthermore,

Table 14 reports a T-test comparison between those with and those without health and safety training and their performance on the *SBT score*. Inspection of Table 14 shows the mean difference in performance trending in the right direction to display *SBT score* construct validity evidence; those who had had health and safety training performed better on average. However, this difference did not reach significance.

Table 15 displays the results of the correlational analysis between the *SBT score* and two measures of health and safety training. Inspection of Table 15 reveals no significant correlations between the number of different health and safety trainings experienced or the number of hours of health and safety training completed and performance on the *SBT score*. However, these correlations are approaching significance, particularly for the relationship between different health and safety training and *SBT score* performance. Together, Table 14 and Table 15's results do not present conclusive construct evidence, however, result patterns are reflective of the directional relationships you would expect; the more health and safety knowledge one has, the safer one would be expected to perform on a safety performance measure.

Table 14

T-test and Comparison between Those With and Without Health and Safety Training and Performance on the SBT Score Measure

Variable	Had never had health and safety training Mean (*SD) n=21*	Had health and safety training Mean (*SD) n=79*	*t* (98)	*p*
Health and safety training	7.3 (.64)	8.24 (1.12)	-1.36	.17

Table 15

Pearson Correlation Analysis between SBT Measures and SBT Participant's Health and Safety Training

	Health and Safety	
SBT Measures	Number of different training programmes completed *n=70*	Hours of training *n=69*
SBT score	.21±	.12

Note. While 79 SBT participants indicated they had health and safety training, 70 specified the number of different health and safety training, 69 specified how many hours, 9 did not provide the number of trainings, and 10 did not provide information on how many hours - hence *n*'s are different in correlation to the T-test table.

± p=.07

Person Impact Variables

When validating any measure, in particular a new measure, it is recognised that there may be factors, such as individual characteristics, which are not related to the construct being measured, yet negatively influence a person's score and can be identified under the category of an *adverse impact* variable. An adverse impact is defined as something which works to the disadvantage of a race, sex, ethnic, or any other type of group (Biddle, 2005). Thomas (2018) compared person variables of age, gender, work experience, computer game experience, forklift experience, and perceived job risk to SBT performance. Thomas (2018) found computer game playing may adversely impact SBT performance, reporting those with computer game experience performed on average better on the SBT, than those without computer game playing experience (see appendix I). In consideration of this finding, the continuous variable of months of computer game experience will be controlled for when examining the SBT's criterion-related validity. Table 16 displays the descriptive information of this variable. Specifically, Table 16's analysis focuses on the current study's ability to accurately control for this composite variable through investigation into its range, skew, and kurtosis. Examination of Table 16 reveals no range restrictions, relatively normal skew, and kurtosis.

Table 16

Descriptive Statistics for SBT Participant Months of Computer Game Playing

Values	*n*	*M* *(SD)*	Minimum	Maximum	Skew	Kurtosis
Months spent playing computer games	32	177.96 (130.31)	19	480	.61	-.51

Criterion-Related Validation Analysis of the SBT

To validate the SBT, the SBT *score* was correlated with the acquaintance safety behaviour criterion measures of safety compliance, participation, voicing, consciousness, risk taking, and rule bending. Table 17 displays the results of the correlation analysis between the *SBT score* and the criterion variables. Validity was investigated separately by two correlational analyses: correlating the *SBT score* with safety behaviour criteria, and a partial correlation analysis comparing the *SBT score* with safety behaviour criteria *while controlling* for computer game experience. In the latter case, the sample size dramatically decreased due to the small number of data points containing information on the months spent playing computer games item.

Examination of Table 17 shows a number of significant correlations between the *SBT score* and measures of safety behaviours. When correlating the SBT with criterion safety behaviour measures using the complete acquaintance sample, a significant relationship is reported between a high score on the SBT and lower score on rule bending behaviours.

After controlling for computer game experience through a partial correlation, SBT criterion-related evidence significantly increased. The relationship between the SBT and rule bending increases in strength and significance. A negative relationship with risk-taking becomes significant. In addition to which, two positive relationships between the SBT and safety compliance and consciousness become significant. Together these results provide criterion-related evidence for the SBT as a measure of these safety behaviours.

Table 17

Correlational Analysis between the SBT Score Measurement and Acquaintance Safety Measures.

	SBT Score	
Safety behaviour measures	Acquaintances *n*=100	Controlling for months spent playing computer games *n*= 30
Safety compliance	.13	.42*
Safety participation	.00	.22
Safety voicing	.02	.25
Safety consciousness	.08	.42*
Risk-taking	-.08	-.41*
Rule bending	-.20*	-.46**

*p<.05. **p<.01.

Does Safety Behaviour Influence Safety Outcomes?

To validate the SBT the assumption was made that the safety behaviour criterion was predictive of safety performance behaviours. Should this assumption be correct, any criterion-related validity evidence, established from this criterion, is evidence of the SBT's ability to measure and predict safety behaviours. To investigate this assumption a comparison was made between the criterion safety behaviour measures and SBT participants' reported near miss, minor injury, and loss of time injuries.

Table 18 provides the descriptive statistics of SBT participants' self-reports on near misses, minor injuries, and time off work accidents. Self-reports were used as an individual is best able to provide minor accident and near miss information. The factual nature of listing accident reports substantially reduces the likelihood of these reports to be influenced by social desirability and impression management biases (Alliger, Lilienfeld, & Mitchell, 1995). Table 19 shows the results of a correlational analysis between safety behaviour measures and the recorded SBT participant near-misses, minor injuries, and loss of time. Examination of the table shows significant negative correlations between acquaintance measures of safety voicing and safety consciousness, and SBT participants' number of medical injuries. This is in addition to a significant positive relationship between risk-taking behaviours and the number of near misses. These relationships support the assumption made that the safety behaviour measure criteria is reflective of safety outcomes. The significant negative relationship between safety voicing and safety consciousness and medical injuries shows those who had greater voicing and consciousness criterion measures had experienced less medical injury incidents. The positive relationship between higher risk-taking criterion scores and near-misses show those who had greater risk-taking behaviours were involved in more near-misses. The small size of the reported correlations reflect the influence of other factors such as unsafe behaviours of others and unsafe environments in near-misses and accidents. These support the previously made assumption that the criteria standard developed by the acquaintances is an accurate reflection of SBT participants' safety behaviours.

Table 18

Descriptive Statistics of SBT Participant Reports of Summed Near-Misses, Minor Injuries, and Loss of Time Injuries.

Values	*n*	Minimum	Maximum	*M* (*SD*)
Near-misses	100	0	1800	7.03 (13.43)
Injuries requiring medical treatment	100	0	100	62.98 (228.93)
Loss of time injuries	100	0	16	1.94 (3.25)

Table 19

Correlation of SBT Participant Near-miss and Workplace Accident Data and Safety Behaviour Measures.

Safety behaviour measures	Near-miss	Medical injury	Loss of time
Safety compliance	.07	-.01	-.03
Safety participation	.05	-.12	-.07
Safety voicing	.02	-.19*	-.12
Safety consciousness	-.08	-.23*	-.17
Risk-taking	.25*	.14	.05
Rule bending	.04	-.02	.00

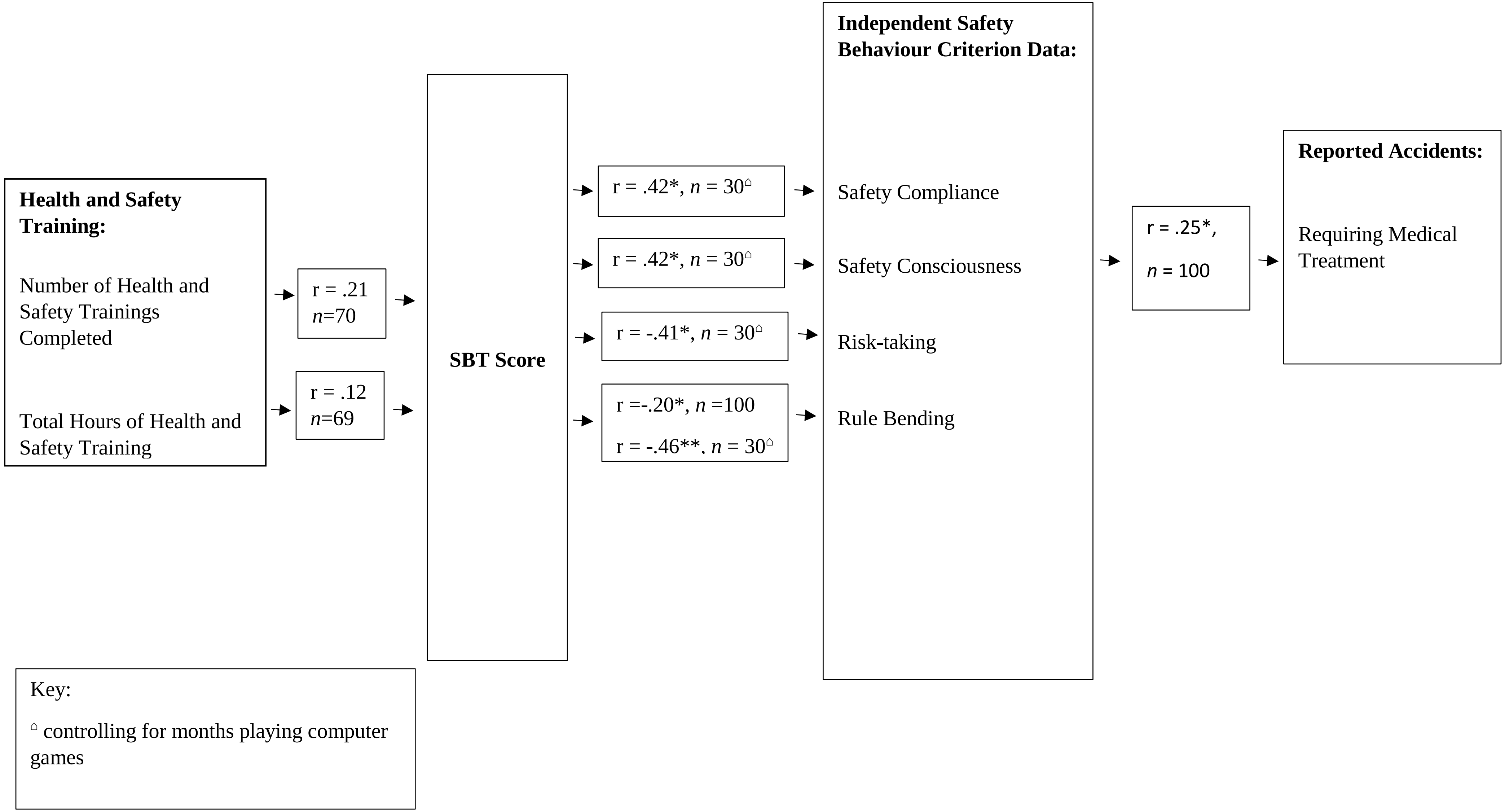
*p<.05.

Results Summary

Figure 4 provides a visual summary of all relevant results.

Figure 4

Summary of SBT Measure Results



General Discussion

Study Summary

The aim of this research was to investigate whether the SBT had criterion-related validity. The criterion-related validity process involved analysis into the associations between performance on the SBT and independent reports on safety behaviour measures. This validation process was conducted to meet the need for an objective and unbiased measure of safety behaviour that can be used during the recruitment process to accurately select an applicant with a greater likelihood of demonstrating safety behaviours.

Summary of Findings

Before examining for measure validity evidence, the identified need that prompted the development of the SBT was investigated. The assumption made by the current study is that current self-report safety behaviour measures are biased by social desirability and impression management processes. Results were consistent with this assumption; self-report measures were inflated by SBT participants in comparison to the independent source. A series of analyses were then run on the SBT to establish evidence of whether it was displaying the anticipated characteristics of a measure. A visual summary of these investigations is displayed in Figure 4. What follows is a brief summary of these findings in sequential order.

Before establishment of the SBT's validity, investigations recognised the safety behaviour criterion used to validate the SBT was likely to be accurate. Investigations into the SBT measure identified that it displayed great measurement characteristics, with normally distributed scores. Although construct validation results trended in the predicted direction, whereby more health and safety training experience was associated with greater SBT performance, no significant evidence was found (see Figure 4). These initial SBT measure investigations lead to the establishment of SBT criterion-related validity evidence and the answering of the research question. As shown in Figure 4, the SBT showed extremely good criterion-related validity results, with the caveat of controlling for participant's months playing computer game experience. Criterion-related validity evidence was found for the SBT as a measure of safety compliance, consciousness, risk-taking, and rule bending. Finally, the safety behaviour criteria was identified as an accurate predictor of safety outcomes (see Figure 4).

So What Does This All Mean?

The SBT was developed to help avoid bias associated with self-report measures of health and safety. Using the assumption that behaviour within a gamified work environment simulation would accurately reflect behaviour in a 'real' situation, the SBT was designed to include decision points which could be scored as either safe or unsafe (risky). The current study's research question "Does the SBT have criterion-related validity?" looked to establish if the SBT met this assumption. If met, individuals with a propensity to behave safely should score higher on the SBT. The correlations with the independent rating of safety behaviour shown in Figure 4 are consistent with the SBT being able to achieve this objective, on the condition that computer game experience is controlled for. Those without computer game experience are being adversely impacted to the extent that the SBT is less able to provide a criterion valid measure of their safety behaviours. This is supported by the fact that SBT participants without computer game experience, on average, performed worse than those with experience (see Appendix I) (Thomas, 2018).

The effect of computer game experience on SBT performance reveals two noteworthy points for the SBT as a criterion valid measure of safety behaviours. (1)

Individuals without computer game experience have a disadvantage over those with computer game experience, and (2) without adjusting for computer game experience, a criterion valid SBT measure of those with no computer game experience is not currently possible.

Current Study Findings and Other Research

To establish the practical feasibility of the SBT, the SBT's criterion-related validity evidence was compared to other commonly used selection tools. In comparison, it was evident that the SBT reported similar relationships. Schmidt & Hunter's (1998) meta-analysis of 19 different selection tools and their relationships with performance criterion variables reported relationships ranging from r=.10 for training and experience to r=.54 for work-samples. Comparatively, when controlling for computer game experience, the current study reported relationships ranging from r=.41 to r=.46. These results suggest the SBT, as a gamified test, offers a predictive power on-par if not better than some currently available selection tools. A proposition not unsupported by gamification research. Gamified tests' ability to provide greater measurement than currently available selection tools was first mentioned in the introduction (section: Benefits of Gamified Assessments) and includes benefits of: measuring within a more authentic context (Shavelson, Baxter, & Pine, 1991), with multiple measurement points (Jaffal & Wloka, 2015), while supporting motivation (Campbell *et al,* 1993; Gangadharbatla & Davis, 2016) in increasing measurement accuracy.

Computer game experience influencing SBT performance is consistent with other gamified-test research. Kim and Shute (2015) examined the influence of computer game experience on participants' performance on their gamified physics test and reported performance on the assessment was influenced by game playing experience.

Participants identified as gamers were found to have had an advantage over non-gamers in achieving test points. Given the novelty of gamified tests, no conclusive research, on why game experience impacts test performance and test measurement, has been identified yet.

In completing the SBT, it is possible that non-computer game players had to focus on learning how to control the game/test and therefore made decision errors due to their unfamiliarity with the test mode, consequently reducing their SBT score. Gameplay research has identified that game playing leads to the development of skills such as computer literacy, spatial abilities, and cognitive attentional skills (Green & Bavelier, 2012), making it possible that game unfamiliarity is limiting game performance. As such, SBT behaviour, as reflected in their SBT score, would not be expected to be associated with their acquaintance's ratings of their safety behaviour.

It is also possible that there is a cognitive aspect associated with computer game playing that is related to behaving safely. Thus, non-game players' SBT performance may have less to do with game control familiarity, and instead be a reflection of poorer cognitive flexibility. Doolittle (1995) reported that playing computer games facilitates cognitive flexibility. He found that college students who played computer games and solved computer riddles were more creative and able to generate alternative hypotheses in a variety of situations. Higher levels of cognitive flexibility has been shown to be associated with lower rates of accidents (Gottfredson, 2004). Gottfredson (2004) reported the information processing skills of being able to quickly identify problematic situations, recall relevant information, and react quickly to unforeseen situations are key skills in accident prevention. It is evidence that a number of explanations could be used for the influence of computer game experience on SBT performance, and its influence needs to be explored in future research.

Practical and Theoretical Implications

Given the unfavourable statistics surrounding workplace accidents and fatalities, and the impact of employee safety behaviours in these accidents, it is important for organisations to hire personnel with the required safety behaviours in an attempt to reduce these unwanted outcomes. The results of this study suggest that, when controlling for game experience, the SBT is a criterion-related valid measure. The practical implications of this finding suggest the SBT could be the solution to the problematic nature of current psychometric safety measures used during the selection process.

However, as noted above, at present SBT scores are adversely impacted by a lack of computer game playing experience. The influence of test mode unfamiliarity should be able to be easily addressed through the use of sufficient pre-test instruction. The SBT instructions shown in Figure 2, while assumed to be sufficient, were not. The static nature of the instructions is likely a hindering factor. Given, the operational aspects of the test mode, further work on enhancing the instructions (Figure 4) using brief animation clips which include point and click control trials, similar to those used in the SBT simulation environment, need to be developed.

Currently, organisations are utilising psychometric tools (e.g. HSI) that measure less influential factors and are subject to biases, such as impression management and social desirability. Job applicants understand that their responses to sensitive issues, such as occupational health and safety, will have a direct impact on how they are perceived by their prospective employer, and consequently adjust their response to suit. As such, there is a strong need for an objective measure that can accurately capture a job applicant’s safety behaviour. The SBT can meet this practical need and provides a criterion-related valid measure, which due to its stealth form of measurement, may offer a more impartial safety behaviour measurement that is not susceptible to social desirability or impression management biases.

The current study's findings also have further, broader, theoretical implications to the future development of gamified assessments. Game experience impacting on gamified tests' criterion-related validity is a novel addition to the small, but growing, body of research into gamified-test design. As outlined in the introduction (section: Gamification Validity), gamified research contains many conflicting reports on the ability to produce valid gamified measures. Some were able to produce valid gamified tests (e.g. Burt, 2017; Mislevy, Almond & Lukas, 2003 Kim & Shute, 2015), while others weren't (e.g. Jaffal & Wloka, 2015; Whetzel *et al*, 2012). Investigations have identified gamified-test design components which impact the validity of these gamified tests include: the complexity and number of constructs for measurement, the construct's translation into gamified measurement items, and gameplay linearity (see Jaffal & Wloka, 2015; Kim & Shute, 2015). Gameplay linearity, for example, has been identified to increase construct validity as every participant is forced into experiencing the same measurement items (Kim & Shute, 2015). The current study's findings contribute to this body of research by identifying that game experience influences gamified test validity, and as such, future studies should consider their participant pool's game experience when establishing gamified measure validity.

Limitations

A limitation of the current study is the number of participants who had computer game experience. Of the SBT's 100 participation sample only 36 had played computer games before completing the SBT. Although the distribution of computer game experience was not too extreme, yes = 36 and no = 62, when attempting to control for computer game experience as a measure of months played, the number dropped down to 30 participants (given not all participants answered every computer experience item). However, it is worth noting that validation research with sample sizes around the 30 mark is not that uncommon. For example, validating research of the Wonderlic Personnel Test included sample sizes of n=20, n=*30*, and n=33 (Wonderlic, 1992).

Two potential theoretical limitations of the criteria safety behaviour measures, is that SBT participants selected their own acquaintance, and that this acquaintance responded using previously validated self-report measures. The benefit of gathering criterion data through acquaintances is that they have no knowledge of the SBT participant's performance on the SBT test, and it thus keeps the criterion standard unbiased from the SBT performance (Cascio & Aguinis, 2005). As a consequence, however, the aforementioned limitations arise and provide new opportunities for error to enter the criterion.

The SBT participant selected their own acquaintance, as only they knew who would be able to report on their safety behaviours. However, given the socially desirable nature of health and safety, it is possible that SBT participants selected an acquaintance who they knew would report favourably on them. The current study attempted to address this by providing an envelope in which all acquaintances were to seal their completed answers in, such that the SBT participant would not have access to them. SBT criterion investigations revealed this potential limitation *did not* impact the safety behaviour criteria. As evident in Table 9, self-report measures were inflated by SBT participants in comparison to their independent source.

The second potentially limiting factor for the safety behaviour criteria was the altering of the previously validated self-report safety behaviour measures. For acquaintances to report on SBT participant's safety behaviours, wording of measurement items was adjusted from a first-person format to suit third-person acquaintance reports on an individual. For example, "*I always use all the necessary safety equipment to do my job"* was adjusted to *"*…always uses….",* where * referred to the individual that asked them to complete the questionnaire. It is possible that by making these changes, error was introduced into the safety behaviour criterion. Research has identified, however, that self-report scales are able to provide a valid and accurate measure when completed by one individual about another (Letzring, 2005; McCrae & Costa, 1987). In addition to which, the current study also reported the safety behaviour criterion is related to safety performance outcomes such as near misses (Figure 4). Thus, neither SBT participants picking their own acquaintances nor acquaintances' use of previous self-report measures appears to have impacted criterion measure accuracy.

A practical limitation of the study concerns the intended use of the SBT. The intended use of the SBT is for high-stake organisational selection decisions, particularly for high-risk positions where safety is paramount. However, the tool was validated using a sample with a range of job risk levels in a low-stakes experimental environment. In this low-stakes environment participants completed the task with no real-life implications or consequences of their performance. In an attempt to address this limitation, SBT participants were informed of the following before completing the test; *"Please imagine that you have applied for a job. The test you are about to complete is being used to determine your suitability for the job. As a job applicant, try to do your best in the test"*. However, it could be argued that results may be different if the tool was implemented on a sample during the recruitment phase of the selection process, where consequences of a poor performance could lead to the applicant being unsuccessful in being hired for a position. The diverse job-risk sample, however, does not appear to have impacted criterion-related validity evidence given there was no reported difference in SBT performance between those who perceive their job to be within a high-risk industry compared to those who perceived their job to be of low-risk (see appendix I).

Future Research

Given the recognised impact of game experience on gamified tests by the current and other studies (e.g. Kim & Shute, 2015), further research into game experience is suggested. In relation to the SBT, as suggested above, further research should look into potentially adding the opportunity for individuals to have a trial-run before they begin the test. The trial run should be on a simplified 'course' that does not include hazards, thus allowing participants to get used to gameplay before measurement without giving participants an advantage. Following such changes, it would be worth then running another criterion-related evidence study, accounting for different game playing experience levels with a larger sample size.

Further research into SBT construct validity is needed. You cannot have criterion-related validity without any evidence of construct validity, however the current study did not report any significant construct evidence. The current study reported correlations between the SBT measure and SBT participants' health and safety training, in part, due to the information being accessible. However, very limited information was available to investigate for construct validity. For example, no information was available on what type of health and safety training was undertaken, nor how long ago the training was completed. Given the necessity for construct evidence, a more in-depth investigation into the SBT's construct evidence is recommended for future research.

The current study was an initial investigation into the SBT and its potential to be used as a gamified psychometric measure of safety behaviour. Before an SBT manual can be developed and the SBT used as a psychometric test, further research into its validity and reliability is also needed. Although the current study reported that the SBT has criterion-related validity, a pattern of validity evidence is required for psychometric tools. To achieve this, research into the SBT's content validity, predictive validity, standard error of measurement, and reliability is recommended. Research into SBT reliability and practice effects are of particular interest given that previous gamified test research has reported practice effects in gamified tests. Performance has been shown to improve after every test completion (e.g. Jaffal & Wloka, 2015). The expectation of psychometric tools to be reliable means the SBT can have no such practice effects, and must provide a reliable measure (American Psychological Association *et al,* 1999).

Further research is also suggested for the gamified-test research field as a whole. Gamified tests are increasingly being designed for high-stake contexts, and as a result gamified test's face validity (how it is perceived as a measure) is becoming increasing important. In the high-stake selection setting for example, applicants who perceive selection systems as unfair have increased test anxiety, and decreased test motivation (Hausknect, Day, & Thomas, 2004), both of which can potentially skew test results (Gangadharbatla & Davis, 2016). Moreover, there is an increased probability of litigation against the organisation (Bauer, Truxillo, Sanchez, Craig, Ferrera, & Campion, 2001). Given this and the novelty of gamification's use as a selection tool, research into perceptions of procedural justice in a high-stakes context is a vital area of exploration for future research.

Conclusion

The evidence from this study suggests that, while there are a number of issues that need to be addressed in future research, the SBT has the potential to be a useful tool for employee selection decisions. The merging of traditional psychometrics, virtual game technology, and test-gamification provides a measurement tool that might be less susceptible to bias than self-report measures. Organisations could benefit from using this tool to objectively measure safety behaviours, thereby potentially reducing workplace accidents and injuries through the identification of job applicants with safe workplace behaviours.

References

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational, & Psychological Testing (US). (1999). *Standards for educational and psychological testing*. Amer Educational Research Assn.

Armstrong, M. B., Landers, R. N., & Collmus, A. B. (2016). Gamifying recruitment, selection, training, and performance management: Game-thinking in human resource management. *In Emerging research and trends in gamification* (pp. 140-165). IGI Global.

Aspen Institute. (2007). *Beyond NCLB: Fulfilling the promise to our nation's children*. Washington, DC: Aspen Institute. Chicago

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of management*, *17*(1), 99-120.

Barrett, P. (2010). *A review of commercial tests and academic articles associate with safety in the workplace*. Retrieved from http://www.pbarrett.net/#whitepapers

Bass, B. M. (1957). Faking by sales applicants on a forced choice personality inventory. *Journal of Applied Psychology*, 41, 403-404.

Bauer, T. N., Truxillo, D. M., Sanchez, R. J., Craig, J. M., Ferrara, P., & Campion, M. A. (2001). Applicant reactions to selection: Development of the selection procedural justice scale (SPJS). *Personnel psychology*, *54*(2), 387-419.

Baxter, G. P., & Shavelson, R. J. (1994). Science performance assessments: Benchmarks and surrogates. *International Journal of Educational Research*, 21, 279–298.

Biddle, D. (2006). *Adverse impact and test validation: A practitioner's guide to valid and defensible employment testing*. Gower Publishing, Ltd.

Breslin, F. C., & Smith, P. (2006). Trial by fire: a multivariate examination of the relation between job tenure and work injuries. *Occupational and Environmental Medicine*, *63*(1), 27-32.

Burke, E. (2009). Preserving the integrity of online testing. *Industrial and Organizational Psychology*, *2*(1), 35-38.

Burt, C. (2017) *Hazard Awareness Test (HAT): Description and Validation Manual*.

Burt, C. D. B. (2015). *New employee safety: Risk factors and management strategies*. Cham: Springer.

Burt, C. D. B., Banks, M. & Williams, S, (2014). The safety risks associated with helping others, *Safety Science*, 62, 136-144. *Published on-line 15-SEP-2013, DOI: 10.1016/j.ssci.2013.08.006.*

Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A theory of performance. In N. Schmitt &W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 35–70). San Francisco: Jossey-Bass.

Cascio, W. F., & Aguinis, H. (2005). Applied psychology in human resource management.

Chmiel, N. (2005). Promoting health work: Self-reported minor injuries, work characteristics, and safety behaviour. In C. *Krunka, P. Hoffmann, & A Bussing (Eds.), Change and quality in human service work* (p 277 – 288). Mering: Rainer Hampp Verlag.

Christian, M. S., Bradley, J. C., Wallace, J. C., & Burke, M. J. (2009). Workplace safety: A meta-analysis of the roles of person and situation factors. *Journal of Applied Psychology*, 94(5), 1103-1127.

Clarke, J. (2009). *Studying the potential of virtual performance assessments for measuring student achievement in science*. Paper presented at the Annual Meeting of the American Educational Research Association (AERA), San Diego, CA.

Clarke-Midura, J., & Dede, C. (2010). Assessment, technology, and change. *Journal of Research on Technology in Education*, 42(3), 309-328.

Darejeh, A., & Salim, S. S. (2016). Gamification Solutions to Enhance Software User Engagement—A Systematic Review. *International Journal of Human-Computer Interaction*, 32(8), 613-642.

Darling-Hammond, L. (2010). Performance counts: Assessment systems that support high-quality learning. *Council of Chief State School Officers.*

Darling-Hammond, L., & Pecheone, R. (2010). *Developing an internationally comparable balanced assessment system that supports high-quality learning*. Washington, DC

Gangadharbatla H. & Davis, D.Z. (Ed.) 2016. *Emerging research and trends in gamification*. Information Science Reference.

de Koster, R. B., Stam, D., & Balk, B. M. (2011). Accidents happen: The influence of safety-specific transformational leadership, safety consciousness, and hazard reducing systems on warehouse accidents. *Journal of Operations management*, 29(7), 753-765.

Deterding, S., Sicart, M., Nacke, L., O'Hara, K., & Dixon, D. (2011). Gamification. using game-design elements in non-gaming contexts. In *CHI'11 extended abstracts on human factors in computing systems* (pp. 2425-2428). ACM.

DeVon, H. A., Block, M. E., Moyle-Wright, P., Ernst, D. M., Hayden, S. J., Lazzara, D. J., ... & Kostas-Polston, E. (2007). A psychometric toolbox for testing validity and reliability. *Journal of Nursing scholarship*, 39(2), 155-164.

Doolittle, J. H. (1995). Using riddles and interactive computer games to teach problem-solving skills. *Teaching Of Psychology*, 22(1), 33.

DuBrin, A. J. (2010). *Impression management in the workplace: Research, theory and practice*. Routledge.

Fastame, M. C., & Penna, M. P. (2012). Does social desirability confound the assessment of self-reported measures of well-being and metacognitive efficiency in young and older adults?. *Clinical Gerontologist*, *35*(3), 239-256.

Field, A. (2009). *Discovering statistics using SPSS*. Sage publications.

Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and individual differences*, *7*(3), 385-400.

Gottfredson, L. S. (2004). Life, death, and intelligence. *Journal of Cognitive Education and Psychology*, *4*(1), 23-46.

Green, C. S., & Bavelier, D. (2012). Learning, attentional control, and action video games. *Current biology*, *22*(6).

Halverson, R., & Owen, V. E. (2014). Game-based assessment: an integrated model for capturing evidence of learning in play. *International Journal of Learning Technology*, *9*(2), 111-138.

Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel psychology*, *57*(3), 639-683.

Herrington, J., Reeves, T. C., & Oliver, R. (2009). *A practical guide to authentic e-learning*. Routledge.

Hofmann, D. A., Morgeson, F. P., & Gerras, S. J. (2003). Climate as a moderator of the relationship between leader-member exchange and content specific citizenship: safety climate as an exemplar. *Journal of Applied Psychology*, *88*(1), 170.

Huang, C. Y., Liao, H. Y., & Chang, S. H. (1998). Social desirability and the clinical self-report inventory: Methodological reconsideration. *Journal of Clinical Psychology*, *54*(4), 517-528.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological bulletin*, *96*(1), 72.

Hunter, J. E., Schmidt, F. L., & Le, H. (2006). Implications of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology*, *91*(3), 594.

Eseryel, D., & Ifenthaler, D. (2012). *Assessment in Game-based Learning: Foundations, Innovations and Perspectives*. Springer.

International Labour Organisation, International Labour Office. (2017). *Occupational Safety Statistics.* Retrieved from http://www.ilo.org/global/topics/safety-and-health-at-work/lang--en/index.htm.

Jaffal, Y., & Wloka, D. (2015). Employing game analytics techniques in the psychometric measurement of game-based assessments with dynamic content. *Journal of e-Learning and Knowledge Society*, *11*(3).

Johnson, T., & Fendrich, M. (2005). Modeling sources of self-report bias in a survey of drug use epidemiology. *Annals of epidemiology*, *15*(5), 381-389.

Ketelhut, D. J., Dede, C., Clarke, J., Nelson, B., & Bowman, C. (2008). Studying situated learning in a multi-user virtual environment. *Assessment of problem solving using simulations*, 37-58.

Kim, Y. J., & Shute, V. J. (2015). The interplay of game elements with psychometric qualities, learning, and enjoyment in game-based assessment. *Computers & Education*, *87*, 340-356.

King, M. F., & Bruner, G. C. (2000). Social desirability bias: A neglected aspect of validity testing. *Psychology and Marketing*, *17*(2), 79-103.

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin*, 107, 34-47.

Letzring, T. D., Block, J., & Funder, D. C. (2005). Ego-control and ego-resiliency: Generalization of self-report scales based on personality descriptions from acquaintances, clinicians, and the self. *Journal of research in personality*, *39*(4), 395-422.

Locke, E. A. (1968). Toward a theory of task motivation and incentives. *Organizational Behavior and Human Performance, 3*(2), 157–189.

McCall, B. P., & Horwitz, I. B. (2005). Occupational vehicular accident claims: a workers’ compensation analysis of Oregon truck drivers 1990–1997. *Accident Analysis & Prevention, 37*(4), 767-774.

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of personality and social psychology*, *52*(1), 81.

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). A brief introduction to evidence-centered design. E*TS Research Report Series, 2003*(1).

Mislevy, R. J., Oranje, A., Bauer, M. I., von Davier, A., Hao, J., Corrigan, S., ... & John, M. (2014). Psychometric considerations in game-based assessment. *GlassLab Report*.

Morassaei, S., Breslin, F. C., Shen, M., & Smith, P. M. (2013). Examining job tenure and lost-time claim rates in Ontario, Canada, over a 10-year period, 1999–2008. *Occup Environ Med*, *70*(3), 171-178.

Neal, A., & Griffin, M. A. (2004). Safety climate and safety at work. *The psychology of workplace safety*, 15-34.

Neal, A., & Griffin, M. A. (2006). A study of the lagged relationships among safety climate, safety motivation, safety behavior, and accidents at the individual and group levels. *Journal of applied psychology*, *91*(4), 946.

Neal, A., Griffin, M. A., & Hart, P. M. (2000). The impact of organizational climate on safety climate and individual behavior. *Safety science*, *34*(1), 99-109.

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European journal of social psychology, 15*(3), 263-280.

OPRA Group. (2017). *OPRA Consulting Group*. Retrieved 28 March, 2017, from http://opragroup.com

Orlitzky, M., & Benjamin, J. D. (2001). Corporate social performance and firm risk: A meta-analytic review. *Business & Society*, *40*(4), 369-396.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of personality and social psychology, 46*(3), 598.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of applied psychology*, *88*(5), 879.

Quellmalz, E. S., & Haertel, G. D. (2004). Use of technology-supported tools for large-scale science assessment: Implications for assessment practice and policy at the state level. *Washington, DC: Center for Education, National Research Council*.

Ramsey, J. D. (1985). Assessment of warnings based on an ergonomic accident sequence model. *international Journal of industrial Ergonomics*, *4*(3), 195-199.

Ramsey, P. (1980). *Ethics at the edges of life: Medical and legal intersections* (No. 1975). Yale University Press.

Reason, J. (1990). *Human error*. Cambridge university press.

Reilly RR, Warech MA. (1993). The validity and fairness of alternatives to cognitive ability tests. In *Wing L, Gifford B (Eds.), Policy issues in employment testing. Boston: Kluwer.*

Reiners, T., & Wood, L. C. (Eds.). (2014). *Gamification in education and business*. Springer.

Resnick, L. B., & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B*. Gifford & M. O'Connor (Eds.), Changing assessments: Alternative views of aptitude*, *achievement*, *and instruction* (pp. 37–75). Norwell, MA: Kluwer Academic Publishers.

Root, N., & Hoefer, M. (1979). The first work-injury data available from new BLS study. *Monthly Lab. Rev.*, *102*, 76.

Rosenquist, A., Shavelson, R. J., & Ruiz-Primo, M. A. (2000). On the "exchangeability" of hands-on and computer simulation science performance assessments. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing, Center for the Study of Evaluation, UCLA.

Roth, P. L., Bobko, P., & McFARLAND, L. Y. N. N. (2005). A meta‐analysis of work sample test validity: Updating and integrating some classic literature. *Personnel Psychology*, *58*(4), 1009-1037.

Safeguard. (2017). Worksafe New Zealand v Broadspectrum (New Zealand) Ltd (DC, 23/12/16). Retrieved from: http://safeguard.co.nz/databases/modus/sgrelated/emohscc/CASE~NZ~NAT~DC~2016~18593?tid=122516913&sid=dd0hql2gs2pa5drcqd71v6namxuss6pb&hli=0&si=1878974471

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological bulletin*, *124*(2), 262.

Shavelson, R. J., Baxter, G. P., & Pine, J. (1991). Performance assessment in science. *Applied measurement in education*, *4*(4), 347-362.

Shute, V. J., Ventura, M., & Kim, Y. J. (2013). Assessment and learning of qualitative physics in newton's playground. *The Journal of Educational Research*, *106*(6), 423-430.

Thomas, K. (2018). Validation of the SBT (Safety Behaviour Test): The Impact of Individual Characteristics on SBT Use and Performance (Unpublished masters's dissertation). University of Canterbury, Christchurch, New Zealand.

Tucker, S., Chmiel, N., Turner, N., Hershcovis, M. S., & Stride, C. B. (2008). Perceived organizational support for safety and employee safety voice: The mediating role of coworker support for safety. *Journal of occupational health psychology*, *13*(4), 319.

Van de Mortel, T. F. (2008). Faking it: social desirability response bias in self-report research. *Australian Journal of Advanced Nursing, The*, *25*(4), 40.

Villanova, P., & Bernardin, H. J. (1991). Performance appraisal: The means, motive, and opportunity to manage impressions. In *R. A. Giacalone & P. Rosenfeld (Eds.), Applied impression management: How image-making affects managerial decisions* (pp. 81-96). Beverly Hills, CA: Sage.

Waltz, C. F., Strickland, O. L., & Lenz, E. R. (Eds.). (2010). *Measurement in nursing and health research*. Springer Publishing Company.

Werbach, K., & Hunter, D. (2012). *For the win: How game thinking can revolutionize your business*. Wharton Digital Press.

Westaby, J. D., & Lee, B. C. (2003). Antecedents of injury among youth in agricultural settings: A longitudinal examination of safety consciousness, dangerous risk taking, and safety knowledge. *Journal of safety research*, *34*(3), 227-240.

Whetzel, D.L., McDaniel, M.A., & Pollack, J.M. (2012). *Work simulations. The handbook of work analysis: Methods, systems, applications and science of work measurement in organizations*, 401-418.

Wonderlic. (1992). Wonderlic personnel test and scholoastic level exam. Milwaukee Avenue, Libertyville.

Worksafe New Zealand. (2016). *Statistics*. Retrieved from http://www.worksafe.govt.nz/worksafe/research/health-and-safety-data.

Appendix A

Acquaintance Questionnaire

UC
UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Code: _______________

General questions about you:

Your Age _____

Your Gender: _______________

How long have you known *… for? Years _____ Months_____

How do you know *… ? (tick as many categories as necessary)

I am *…'s Work colleague □
Work manager□
Worker supervisor□
Spouse □
Child □
Parent □
Friend □
Sport/Recreation associate□
Partner □
Other □ (please specify_____________________________)

Please indicate how well you know *…. by placing a mark on the 100 point scale.

0.....10…..20…..30…..40…..50…..60…..70…..80…..90…..100

Not Very Well At All Extremely Well

Questions about *…:

Rated safety risk score

Considering *…'s behaviour in all the situations that you know, please indicate *…'s general degree of safety risk by placing a mark on the 100 point scale.

0.....10…..20…..30…..40…..50…..60…..70…..80…..90…..100

Not at all Risky Extremely Risky

For each of the three accident and incidents categories please indicate the number you know *… has had in each of the three locations: at work, at home, in any other location. If you feel you don't know please tick in the 'don't know' column

Accident/incident Category	At work	At home	Other Location (e.g., while on holiday, recreating)	Don't Know
Near miss incident, which could have resulted in injury or damage				
Minor injury requiring medical attention (e.g. first aid treatment or a visit to a doctor)				
Lost Time Injury (LTI) that has required time off work				

Safety Compliance and Participation

These statements are about how *… behaves. For each statement, please circle the number which indicates the extent to which you disagree or agree. If you don't know about an item please tick in the 'don't know' column.

	Strongly disagree	Disagree	Neither agree/disagree	Agree	Strongly Agree
*… always uses all the necessary safety equipment to do their job	1	2	3	4	5
*… always uses the correct safety procedures to carry out their job	1	2	3	4	5
*… always ensures the highest level of safety to carry out their job	1	2	3	4	5
*…. promotes the safety programme within their organisation	1	2	3	4	5
*… puts in extra effort to improve the safety of their workplace	1	2	3	4	5
*… voluntarily carries out tasks or activities that help to improve workplace safety	1	2	3	4	5

Safety Voicing

Listed below are a number of statements that could be used to describe *…'s safety behaviour. Please circle a number to indicate how much you agree or disagree with each statement. If you don't know about any item please tick in the 'don't know' column.

	Strongly disagree	Disagree	Neither agree/disagree	Agree	Strongly Agree
*… makes suggestions about how safety could be improved	1	2	3	4	5

*… tells others who are doing something unsafe to stop	1	2	3	4	5
*… discuss new ways to improve safety with his/her colleagues or boss	1	2	3	4	5
*… informs the boss when he/she notices a potential hazard	1	2	3	4	5
*… reports to his/her boss if their colleagues break any safety rules	1	2	3	4	5

<u>Rule Bending</u>

Listed below are a number of statements that could be used to describe *…'s safety behaviour. Please circle a number to indicate how much you agree or disagree with each statement. If you don't know about any item please tick in the 'don't know' column.

	Strongly disagree	Disagree	Neither agree/disagree	Agree	Strongly Agree
*… sometimes cuts corners if it makes the task easier	1	2	3	4	5
Work pressures means that *… sometimes bends the rules	1	2	3	4	5
Occasionally*….bends the rules when he/she knows it is safe to do so	1	2	3	4	5
When *…'s boss is not around he/she can be more flexible with which procedures he/she follows	1	2	3	4	5

<u>Safety Consciousness and Risk Taking</u>

These statements are about how *… behaves. For each statement, please circle the number which indicates the extent to which you disagree or agree. If you don't know about an item please tick in the 'don't know' column.

	Strongly disagree	Disagree	Neither agree/disagree	Agree	Strongly Agree
*… always takes extra time to do things safely	1	2	3	4	5
People think of *… as being an extremely safety-minded person	1	2	3	4	5
*… always avoids dangerous situations	1	2	3	4	5
*… takes a lot of time to do things safely even when it slows their performance	1	2	3	4	5
*… often makes sure that other people do things that are safe and healthy	1	2	3	4	5
*… gets upset when seeing other people acting dangerously	1	2	3	4	5
*… thinks doing the safest possible thing is always the best thing	1	2	3	4	5
*… would rather take risks than be overly cautious	1	2	3	4	5
In the past month *… has done some exciting things that other people might think are dangerous	1	2	3	4	5
*… loves to take risks even when there is a small chance he/she could get hurt	1	2	3	4	5
Sometimes people get on *… nerves when they tell	1	2	3	4	5

him/her how to act “more safely”					
*… values having fun more than being safe	1	2	3	4	5

Please check that you have answered all questions.

Thank you for taking the time to participate in this research.

Appendix B

Individual Characteristics Questionnaire

Code:_________

Demographic Questions

- Age:……….   - Gender:……….

Computer Game Experience

- Have you played a computer game before using the CPT?

  ☐ Yes   ☐ No (please go to "forklift use experience" questions)

- Please indicate how many years and months you have been playing computer games.

  ……….Years and……….months

- Have you ever played a point and click game?

  ☐ Yes   ☐ No   ☐ Don't know

- How often do you play computer games?

  ☐ Daily   ☐ Once every 6 months
  ☐ Weekly   ☐ Once a year
  ☐ Monthly   ☐ Less than once a year

Forklift use experience

- Have you ever driven a forklift?

☐ Yes ☐ No (please go to “work experience” questions)

- Do you have a forklift licence?

  ☐ Yes ☐ No

- Please indicate how many years and months you have had a forklift licence for.

  ……….Years and……….Months

- How many jobs have you worked in that have required you to use a forklift? ………

- Please indicate how many years and months you have worked in jobs that have required you to use a forklift.

  ……….Years and……….months

- How many hours of forklift training have you had?..........

Work Experience

- Do you work full time or part time? Full time □ Part time □

- Please indicate how many years and months you have had your current job for.

  ……….Years and……….months

- How many co-workers do you currently have?..........

- In total, how many different jobs have you had?..........

- Please indicate how many years and months you have worked for, in total.

  ……….Years and……….months

- How many different organisations have you worked for?..........

Accidents and Incidents

- For each of the three accident and incidents categories please indicate the number you have had in each of the three locations: at work, at home, other.

Accident/Incident Category	At work	At home	Other location (e.g. while on holiday, recreating)
Near miss incidents, which could have resulted in injury or damage			
M**inor injury** requiring medical attention (e.g. first aid treatment or a visit to a doctor)			
Lost Time Injury (LTI) that has required time off work			

Perceived Job Safety Risk

Please indicate the safety risk associated with your current job by placing a mark on this 100 point scale.

0.....10…..20…..30…..40…..50…..60…..70…..80…..90…..100

Not at all Risky Extremely Risky

Health and Safety Training

- Have you completed any health and safety training?

☐ Yes ☐ No (please go to “CPT Usability” questions)

- How many different training programmes related to health and safety have you completed?.............

- How many hours of training related to health and safety have you completed?……….

CPT (Compliance and Participation Test) Usability: The following questions are about your experience with the test you just completed.

- How understandable were the instructions given to you to use the CPT (please circle a number)?

  1……….2………..3………..4………..5……….6………..7

  Not at all Completely

- How easy was it to control the forklift in the CPT (please circle a number)?

  1……….2………..3………..4………..5……….6………..7

  Very Hard Very Easy

- How appropriate was the speed that the forklift moved in the CPT?

  1……….2………..3………..4………..5……….6………..7

  Very Inappropriate Very Appropriate

- Overall, how easy was it to complete the CPT?

  1……….2………..3………..4………..5……….6………..7

  Very Hard Very Easy

- How much did you enjoy completing the CPT

  1……….2………..3………..4………..5……….6………..7

  Not at all Completely

- Do you have any other comments in regard to using the CPT (please write below)?

____

____

____

Safety Participation and Compliance

Listed below are a number of statements that could be used to describe your safety behaviour. Please circle a number to indicate how much you agree or disagree with each statement.

	Strongly disagree	Disagree	Neither agree/disagree	Agree	Strongly Agree
I use all the necessary safety equipment to do my job	1	2	3	4	5
I use the correct safety procedures for carrying out my job	1	2	3	4	5
I ensure the highest levels of safety when I carry out my job	1	2	3	4	5
I promote the safety program within the organisation	1	2	3	4	5
I put in extra effort to improve the safety of the workplace	1	2	3	4	5
I voluntarily carry out tasks or activities that help to improve workplace health and safety	1	2	3	4	5

Rule bending

Listed below are a number of statements that could be used to describe your safety behaviour. Please circle a number to indicate how much you agree or disagree with each statement.

	Strongly disagree	Disagree	Neither agree/disagree	Agree	Strongly Agree
I cut corners if it makes the task easier	1	2	3	4	5
Work pressures mean that I bend safety rules	1	2	3	4	5
I bend the rules when I know it is safe to do so	1	2	3	4	5
When my boss is not around I can be more flexible with which procedures I follow	1	2	3	4	5

Safety Voicing

Listed below are a number of statements that could be used to describe your safety behaviour. Please circle a number to indicate how much you agree or disagree with each statement.

	Strongly disagree	Disagree	Neither agree/disagree	Agree	Strongly Agree
I make suggestions about how safety could be improved	1	2	3	4	5
I tell colleagues who are doing something unsafe to stop	1	2	3	4	5
I discuss new ways to improve safety with my colleagues or boss	1	2	3	4	5
I inform the boss when I noticed a potential hazard	1	2	3	4	5
I report to my boss if my colleagues break any safety rules	1	2	3	4	5

Safety Consciousness and Risk Taking

Listed below are a number of statements that could be used to describe your safety behaviour. Please circle a number to indicate how much you agree or disagree with each statement.

	Strongly disagree	Disagree	Neither agree/disagree	Agree	Strongly Agree
I always take extra time to do things safely	1	2	3	4	5
People think of me as being an extremely safety-minded person	1	2	3	4	5
I always avoid dangerous situations	1	2	3	4	5
I take a lot of time to something safely even when it slows my performance	1	2	3	4	5
I often find myself making sure that other people do things that are safe and healthy	1	2	3	4	5
I get upset when I see other people acting dangerously	1	2	3	4	5
Doing the safest possible thing is always the best thing	1	2	3	4	5
I would rather take risks than be overly cautious	1	2	3	4	5
In the past month I have done some exciting things that other people might think are dangerous	1	2	3	4	5
I love to take risks even when there is a small chance I could get hurt	1	2	3	4	5
Sometimes people get on my nerves when they tell me how to act “more safely”	1	2	3	4	5
I value having fun more than being safe	1	2	3	4	5

Please check that you have answered all questions.

Thank you for taking the time to participate in this research.

Appendix C

Email Template

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Dear *

My name is Lydia Crowe, and I am completing my Masters at the University of Canterbury. As part of my research, I am validating a new test of employee safety behaviour called the CPT. The CPT (safety compliance and participation test) is gamified test, meaning that people interact with an environment on a computer as a way of being tested. Hopefully the test can be used in organisations to inform which employees require health and safety training, and in the selection process to identify applicants who will behave safely in a work environment.

My Maters research is seeking to validate the CPT. Validation is a three stage process, where participants must first complete the CPT, then fill out a questionnaire regarding their own safety behaviour, and finally have a colleague/associate/supervisor fill out a questionnaire regarding the participant's safety behaviour. In exchange, a $10 petrol voucher will be given to both the participant and the colleague/associate/supervisor. The process requires about 30 minutes of time from the participant, and about 10 minutes of time from the colleague/associate/supervisor. To complete the study I could work at your organisation, and would require a quiet office space and an internet connection.

At the end of the study, if the CPT is shown to be a valid measure of safety behaviour, I will be able to provide your organisation with aggregated information on how your employees did on the test. While I am unable to give you data for individual employees, I can give you a distribution of their scores.

Thank you very much for considering to be part of my study. I would be happy to meet and demonstrate the CPT, and discuss the project further.

For more information, please contact me at lydia.crowe@pg.canterbury.ac.nz.

Kind regards,

Lydia Crowe

Appendix D

Participant Advertisement

Participants Wanted!

Would you like to take part in a study that investigates using a computer game as a new measure of safety behaviour?

I am conducting a study that aims to validate a computer game as a new measure of safety behaviour.

Participants will be required to play the computer game (which will take approximately 20 minutes), complete a questionnaire (which will take approximately 10 minutes), and find an acquaintance to complete a questionnaire as well (which will take them approximately 10 minutes).

The participant and their acquaintance will each receive a $10 petrol voucher after completing their tasks.

Participants are required to have adequate eye sight for playing a computer game, and to be currently working either full time or part time.

If you are interested, please tear off an email address and contact me!

lydia.crowe@pg.canterbury.
lydia.crowe@pg.canterbury.
lydia.crowe@pg.canterbury.
lydia.crowe@pg.canterbury.
lydia.crowe@pg.canterbury.
lydia.crowe@pg.canterbury.
lydia.crowe@pg.canterbury.
lydia.crowe@pg.canterbury.
lydia.crowe@pg.canterbury.
lydia.crowe@pg.canterbury.
lydia.crowe@pg.canterbury.
lydia.crowe@pg.canterbury.
lydia.crowe@pg.canterbury.
lydia.crowe@pg.canterbury.
lydia.crowe@pg.canterbury.
lydia.crowe@pg.canterbury.
lydia.crowe@pg.canterbury.
lydia.crowe@pg.canterbury.

Appendix E

CPT Participant Information Sheet

Department of Psychology
Email: kth63@uclive.ac.nz

05/04/2017

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Validation of the CPT: The Impact of Individual Characteristics on CPT
Participant Information Sheet

I am Kristy Thomas, and I am a Masters student in the Department of Psychology at the University of Canterbury. The purpose of this research is to validate the safety *Compliance and Participation Test* (CPT) as a measure of safety behaviour. The CPT is a fully animated computer game. The players must point the cursor at areas on the screen and click in order to interact with the game environment. In the CPT, players will be given instructions to retrieve several different items from within a warehouse using a forklift, and then load each item into a container. In order to validate the CPT, the current study will require participants to complete both the CPT, and an individual characteristics questionnaire. The results of the questionnaire will be used to determine if any individual characteristics have an identifiable impact on CPT use and performance.

If you choose to take part in this study, your involvement in this project will be to complete the CPT, and to complete a questionnaire that assesses individual characteristics and safety behaviour. The CPT and the questionnaire will each take approximately 20 minutes to complete. You will also be required to invite one of your acquaintances to participant in the study. Your acquaintance can be a work colleague, friend, family member, or sports and recreation associate for example, and must be close enough to you to be able to report on your safety behaviour in general and at work. You will be required to take an envelope to the acquaintance you have chosen, which will contain an information sheet, a consent form for them to sign, a questionnaire for them to complete that assesses your safety behaviour, and a $10 petrol voucher for your acquaintance.

Participation is voluntary and you have the right to withdraw at any stage without penalty. You may ask for your raw data to be returned to you or destroyed at any point. If you withdraw, I will remove information relating to you. However, once analysis of raw data starts on the *1st October 2017*, it will become increasingly difficult to remove the influence of your data on the results.

The results of the project may be published, but you may be assured of the complete confidentiality of data gathered in this investigation: your identity will not be made public without your prior consent. To ensure confidentiality, you will be allocated a code that will be written on your CPT score, your questionnaire, and your acquaintance's questionnaire as opposed to your name. Furthermore, all physical data will be stored in a locked filing cabinet in a locked room, while all electronic data will be stored in a password protected computer in a locked room, and no person outside of the research team will have access to data. A thesis is a public document and will be available through the UCLibrary. Data will be destroyed after five years, unless a publication outlet requires extended archiving of the data.

Please indicate to the researcher on the consent form if you would like to receive a copy of the summary of results of the project.

The project is being carried out in partial fulfilment of the requirements for the degree of Master of Science in Applied Psychology at the University of Canterbury by Kristy Thomas under the supervision of Associate Professor Christopher Burt, who can be contacted at christopher.burt@canterbury.ac.nz. He will be pleased to discuss any concerns you may have about participation in the project.

This project has been reviewed and approved by the University of Canterbury Human Ethics Committee, and participants should address any complaints to The Chair, Human Ethics Committee, University of Canterbury, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz).

If you agree to participate in the study, you are asked to complete the consent form and return to the researcher.

Appendix F

CPT Participant Consent Form

Department of Psychology
Email: kth63@uclive.ac.nz
05/04/2017

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Validation of The CPT: The Impact of Individual Characteristics on CPT Use.

CPT Participant Consent Form

☐ I have been given a full explanation of this project and have had the opportunity to ask questions.

☐ I understand what is required of me if I agree to take part in the research.

☐ I understand that participation is voluntary and I may withdraw at any time without penalty. Withdrawal of participation will also include the withdrawal of any information I have provided should this remain practically achievable.

☐ I understand that any information or opinions I provide will be kept confidential to the researcher and supervisor of the research and that any published or reported results will not identify the participants or organisation. I understand that a thesis is a public document and will be available through the UC Library.

☐ I understand that all data collected for the study will be kept in locked and secure facilities and/or in password protected electronic form and will be destroyed after five years, unless a publication outlet requires extended archiving of the data.

☐ I understand that there are no risks associated with taking part in this study

☐ I understand that I can contact the researcher, Kristy Thomas (kth63@uclive.ac.nz) or supervisor Christopher Burt (christopher.burt@canterbury.ac.nz) for further information. If I have any complaints, I can contact the Chair of the University of Canterbury Human Ethics Committee, Private Bag 4800,Christchurch (human-ethics@canterbury.ac.nz)

☐ I would like a summary of the results of the project.

☐ By signing below, I agree to participate in this research project.

Name:__________________Signed:________________Date:__________

Email address *(for report of findings, if applicable)*:

____

Appendix G

Acquaintance Information Sheet

Department of Applied Psychology
Email: lydia.crowe@pg.canterbury.ac.nz
Date: 04.04.2017

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Validation of the Compliance and Participation Test: Criteria Validity Evidence

Acquaintance Information Sheet

I'm Lydia Crowe and I am a Masters of Applied Psychology student at the University of Canterbury conducting a study of the validity of the *Compliance Participation Test (CPT)*. The purpose of the research is to establish if the CPT is a valid measure of safety compliance and participation.

If you choose to take part in this study, your involvement in this project will be to spend approximately 10minutes completing an acquaintance questionnaire. This questionnaire includes several safety behaviour items about *…………………………..…who consented to you completing this questionnaire. Whenever you see *… below this refers to the person who invited you to participate in this study. After completing this questionnaire, the questionnaire itself and the completed consent form should be sealed in the provided envelope and given back to the demonstrator who will bring it back to University of Canterbury Psychology Department and collect a $10 petrol voucher for your participation in this study.

Participation is voluntary and you have the right to withdraw at any stage without penalty. You may ask for your raw data to be returned to you or destroyed at any point. If you withdraw, I will remove information provided by you. However, once analysis of raw data starts on the 1st October 2017, it will become increasingly difficult to remove the influence of your data on the results.

The results of the project may be published, but you may be assured of the complete confidentiality of data gathered in this investigation: your identity will not be made public without your prior consent. To ensure confidentiality the consent form and questionnaire will be kept in the sealed envelope until its arrival at the University of Canterbury Psychology Department. When on opening the two documents will be immediately separated to maintain confidentiality. Your name will not be collected on any document other than the separate consent form. Instead the Acquaintance Questionnaire will be coded with the same code as *….No one other than me (as the researcher) and Chris Burt(as the research supervisor) will have access to the data. Physical data will be stored in a locked filing cabinet in a locked room. Electronic data will be stored on a password protected computer, in a locked room. Data will be destroyed after 5 years, unless a publication outlet requires extended archiving of the data. A thesis is a public document and the subsequent thesis will be available through the UC Library. Please indicate on the consent form if you would like to receive a copy of the summary of results of the project.

The project is being carried out as a requirement for the Masters of Applied Psychology programme by Lydia Crowe under the supervision of Chris Burt, who can be contacted at christopher.burt@canterbury.ac.nz. He will be pleased to discuss any concerns you may have about participation in the project.

This project has been reviewed and on approved by the University of Canterbury Human Ethics Committee, and participants should address any complaints to The Chair, Human Ethics Committee, University of Canterbury, Private Bag 4800, Christchurch(human-ethics@canterbury.ac.nz). The study's reference number is HEC 2017/26.

If you agree to participate in the study, you are asked to complete the consent form before completing the acquaintance questionnaire.

Appendix H

Acquaintance Consent Form

Department of Applied Psychology
Email: lydia.crowe@pg.canterbury.ac.nz
Date: 04.04.2017

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Validation of the Compliance and Participation Test: Criteria Validity Evidence

Acquaintance Consent Form

☐ I have been given a full explanation of this project and have had the opportunity to ask questions.

☐ I understand what is required of me if I agree to take part in the research.

☐ I understand that participation is voluntary and I may withdraw at any time without penalty. Withdrawal of participation will also include the withdrawal of any information I have provided should this remain practically achievable.

☐ I understand that any information or opinions I provide will be kept confidential to the researcher and that any published or reported results will not identify the participants. I understand that a thesis is a public document and will be available through the UC Library.

☐ I understand that all data collected for the study will be kept in locked and secure facilities and/or in password protected electronic form and will be destroyed after five years, unless a publication outlet requires extended archiving of the data.

☐ I understand there are no risks associated with taking part in this study

☐ I understand that I can contact the researcher [*Lydia Crowe* lydia.crowe@pg.canterbury.ac.nz] or supervisor [*Chris Burt*christopher.burt@canterbury.ac.nz] for further information. If I have any complaints, I can contact the Chair of the University of Canterbury Human Ethics Committee, Private Bag 4800,Christchurch (human-ethics@canterbury.ac.nz)

☐ I would like a summary of the results of the project.

☐ By signing below, I agree to participate in this research project.

Name:___________________________Signed:______________________Date__________

Email address *(for report of findings, if applicable)*:

____

Appendix I

Differences in SBT performance based on game playing experience

Independent sample t-test comparing mean scores on the SBT measures between participants that had never played computer games and participants that had played computer games.

SBT measures	Had never played computer games Mean (SD) N=62	Had played computer games Mean (SD) N=36	*t*-test result (96)
SBT score	7.66 (2.66)	8.91 (2.61)	-2.26*
SBT test time†	1114.85 (178.56)	1018.00 (144.72)	2.75**

†After removing outliers *SBT test time*, N=60 for participants that had never played computer games, and df=94 for the *t*-test.
*p<0.05, **p<0.01

Note: Table taken from Thomas (2018).

Appendix J

Differences in SBT performance based on game playing experience

Pearson correlations between mean scores on the SBT measures and the perceived job risk of participants.

SBT measures	Perceived job risk N= 98
SBT score	.01
SBT test time†	.00

†After removing outliers *SBT test time*, N=96 for *perceived job risk.*

Note: Table taken from Thomas (2018).